

# Refined analysis of sparse MIMO radar

Dominik Dorsch \*

Holger Rauhut<sup>†</sup>

September 14, 2015

## Abstract

We analyze a multiple-input multiple-output (MIMO) radar model and provide recovery results for a compressed sensing (CS) approach. In MIMO radar different pulses are emitted by several transmitters and the echoes are recorded at several receiver nodes. Under reasonable assumptions the transformation from emitted pulses to the received echoes can approximately be regarded as linear. For the considered model, and many radar tasks in general, sparsity of targets within the considered angle-range-Doppler domain is a natural assumption. Therefore, it is possible to apply methods from CS in order to reconstruct the parameters of the targets. Assuming Gaussian random pulses the resulting measurement matrix becomes a highly structured random matrix. Our first main result provides an estimate for the well-known restricted isometry property (RIP) ensuring stable and robust recovery. We require more measurements than standard results from CS, like for example those for Gaussian random measurements. Nevertheless, we show that due to the special structure of the considered measurement matrix our RIP result is in fact optimal (up to possibly logarithmic factors). Our further two main results on nonuniform recovery (i.e., for a fixed sparse target scene) reveal how the fine structure of the support set — not only the size — affects the (nonuniform) recovery performance. We show that for certain “balanced” support sets reconstruction with essentially the optimal number of measurements is possible. Indeed, we introduce a parameter measuring the well-behavedness of the support set and resemble standard results from CS for near-optimal parameter choices. We prove recovery results for both perfect recovery of the support set in case of exactly sparse vectors and an  $\ell_2$ -norm approximation result for reconstruction under sparsity defect. Our analysis complements earlier work by Strohmer & Friedlander and deepens the understanding of the considered MIMO radar model.

**AMS Subject Classification:** 94A20, 94A12, 60B20, 90C25, 65F22,

**Keywords:** MIMO radar, compressed sensing,  $\ell_1$ -minimization, restricted isometry property, LASSO, random matrix

## 1 Introduction

MIMO (multiple-input multiple-output) radar systems can simultaneously transmit several uncorrelated waveforms from spatially distributed transmitters and record the reflected signals at different receiver locations. The transmit/receive antennas may either be widely separated giving the possibility to view the targets from different angles or co-located antenna configurations — to be studied in this paper — can yield superior resolutions and target identifiability as compared to standard phased-array radar [13, 21, 22].

When it comes to the implementation of a MIMO radar system involving several waveforms, one is naturally confronted with an increased amount of data to be processed. Yet in MIMO radar the description of the target parameters can turn out to be particularly sparse, i.e., the target scene can be modeled as a vector of the object reflectivities with the property that “almost” all entries are zero [8]. Since it is often possible to assume the transformation of the transmitted signals through the channel to be approximately linear, one is faced with the reconstruction of the targets’ parameters from highly incomplete linear measurements.

---

\*Chair for Mathematics C (Analysis), RWTH Aachen University, dorsch@mathc.rwth-aachen.de

<sup>†</sup>Chair for Mathematics C (Analysis), RWTH Aachen University, rauhut@mathc.rwth-aachen.de

During the last decade, the theory of compressed sensing (CS) evolved around this particular setting. In CS one typically tries to reconstruct an unknown vector  $\mathbf{x}$  from highly underdetermined linear measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}$  under the additional assumption of sparsity in  $\mathbf{x}$ . While only suboptimal results are available so far for deterministic measurement matrices  $\mathbf{A}$  — see for instance the discussion in [14, Chapter 6.1] — it is rather typical to consider matrices which are defined in terms of random variables. The crucial point is that, apart from theory, also in practice the concept of randomness can be implemented efficiently and has already proven effective. This opens the field for advanced methods from probability theory making it possible to prove strong recovery results — which usually hold true with very high probability. Typical results guaranteeing recovery of sparse vectors  $\mathbf{x}$  from measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}$  provide conditions on the minimal number of required measurements — ideally scaling linearly on the number of nonzero entries in  $\mathbf{x}$  up to logarithmic factors. Given this, the theory of CS provides several algorithms for reconstruction. Usually  $\mathbf{x}$  can either be obtained as the solution to a convex optimization problem, or by an iterative (greedy) algorithm [14].

Ideas from CS have probably first been applied to radar in [1, 17, 12]. The extension to MIMO radar has been conducted in [8, 26, 31, 32]. In a recent paper, Strohmer & Friedlander [27] analyze a particular model of co-located MIMO radar and prove recovery results. They assume random probing signals which leads to a measurement process constituting of  $N_R N_t$  random measurements, where  $N_R$  is the number of receiver nodes and  $N_t$  is the number of (time domain) samples being taken at each receiver. Under the additional assumption of randomly distributed targets, they show that basically the condition  $N_R N_t \gtrsim s \log(N)$ , where  $s$  is the number of targets and  $N$  is the dimension of the target scene  $\mathbf{x}$ , is sufficient for reconstruction by minimizing the so-called LASSO functional. Thus, typical results from CS on random measurements are resembled for these specific MIMO radar measurements.

We adopt this model and deepen its understanding by deriving further properties and results in the context of CS. First, we prove an estimate for the restricted isometry property (RIP) of the involved measurement matrix. The RIP guarantees uniform recovery (i.e., simultaneous recovery of all sufficiently sparse vectors from a single random draw of the matrix with high probability) and, moreover, yields strong error estimates for noisy measurements and only approximately sparse vectors. Compared to standard estimates for standard random matrix constructions, we require more measurements for a given sparsity which, however, as we show, is still optimal and due to the special structure of the considered matrix. Motivated by the somehow better results in [27], we furthermore deepen the analysis of the measurement process by introducing a parameter for the fine structure of the supports sets. In this way we are able to provide nonuniform recovery results for “balanced” support sets resembling the required number of measurements one would usually expect from common CS theory. Ultimately, this explains the result in [27], since random support sets are — on average — sufficiently balanced and, hence, for randomly distributed targets less measurements are sufficient. To the best of our knowledge, it has not been observed earlier for realistic measurement matrices that the recovery performance may depend significantly on the fine structure of the support sets.

## 1.1 The MIMO radar model

Our MIMO radar model consists of  $N_T$  transmit and  $N_R$  receive antennas. We consider the setting from [27] where some assumptions on the geometry of the antenna/target locations have been made. We assume that

- the antenna arrays and the scatterers are located in the same two-dimensional plane,
- the transmit/receive antennas are located along one common line (“monostatic radar”),
- the distance of the targets from the antenna arrays is sufficiently large so that the radar return of any scatterer can be considered to be fully correlated across the array (“coherent propagation scenario”).

We point out that our results can be generalized to a three-dimensional setting. However, since the effects we want to study already occur in two dimensions, we concentrate on this case in order to keep the notation simple. The transmit antennas occupy the positions  $(0, id_T\lambda) \in \mathbb{R}^2$ ,  $i = 0, 1, \dots, N_T - 1$ , where  $\lambda$  denotes the wavelength of the carrier frequency of the radar system. The receive antennas are located at  $(0, jd_R\lambda)$ ,  $j = 0, 1, \dots, N_R - 1$ . It is known that by choosing  $d_T = 1/2$  and  $d_R = N_T/2$  or,

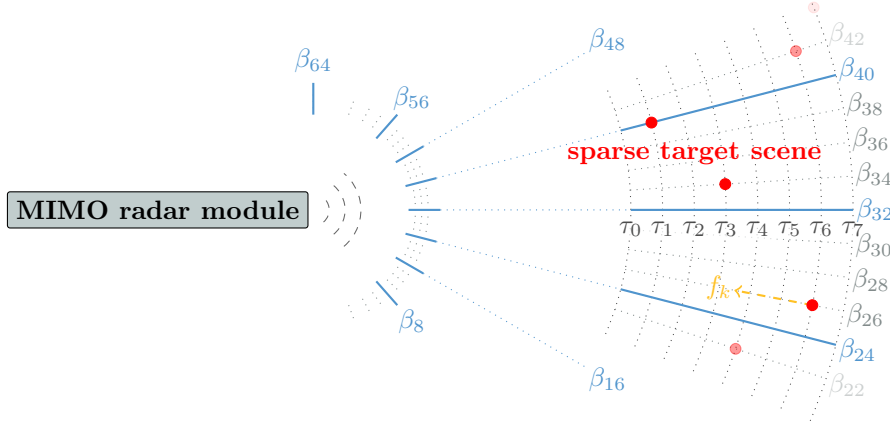


Figure 1: Sparse MIMO radar: sector of the physical target domain for the case of  $N_T = 8$  and  $N_R = 8$  transmit/receive antennas, yielding an angular resolution of  $N_T N_R = 64$ . An exemplary delay grid with a resolution of  $N_\tau = 8$  is depicted. The considered Doppler effect introduces a third dimension, parametrized by the Doppler shifts  $f$ .

alternatively,  $d_T = N_R/2$  and  $d_R = 1/2$  similar characteristics as those of a virtual array with  $N_T N_R$  antennas can be obtained [15]. In this sense these particular choices for  $d_T$  and  $d_R$  are favorable compared to other choices in practice. This fact will also become clear during our analysis. Throughout the paper we concentrate on the case where

$$d_T = \frac{1}{2}, \quad d_R = \frac{N_T}{2}. \quad (1)$$

The second case can be treated analogously.

The  $i$ th transmit antenna repeatedly sends a fixed complex continuous-time signal  $s_i = s_i(t)$  of period duration  $T$ . The reflected signals, due to reflections caused by one unit reflectivity target at position  $(r \cos(\theta), r \sin(\theta)) \in \mathbb{R}^2$ , traveling with radial speed  $v$ , will be recorded at the  $j$ th receiver (cf. [32]) as

$$y_j(t) = \sum_{i=1}^{N_T} e^{-i2\pi \cdot c\lambda^{-1}(t-d_{i,j}(t)/c)} s_i(t - d_{i,j}(t)/c),$$

where  $c$  denotes the speed of light and  $d_{i,j}(t)$  is the distance the emitted wave has to travel from the  $i$ th transmitter to the  $j$ th receiver at time  $t$ , given by

$$d_{i,j}(t) = 2(r + vt) + \sin(\theta)d_T\lambda(i-1) + \sin(\theta)d_R\lambda(j-1).$$

After demodulation (multiplication with  $e^{i2\pi \cdot c\lambda^{-1}t}$ ), assuming narrowband transmit waveforms, slowly moving targets, and a far field scenario, i.e.,  $r \gg \max\{d_T N_T \lambda, d_R N_R \lambda\}$ , the received baseband signal at the  $j$ th receiver is approximately given by

$$y_j(t) \approx e^{i2\pi \cdot 2\lambda^{-1}r} e^{i2\pi \cdot \sin(\theta)d_R(j-1)} \sum_{i=1}^{N_T} e^{i2\pi \cdot 2\lambda^{-1}vt} e^{i2\pi \cdot \sin(\theta)d_T(i-1)} s_i(t - 2r/c). \quad (2)$$

(See also [32], where the antennas are distributed freely over a small area, though.)

The parameters of a target are given by the triple  $(\theta, r, v)$  representing the position (in radial coordinates) and the radial speed. It will, however, be more convenient in the remaining part of this paper to equivalently consider the triples  $(\sin(\theta), 2r/c, 2\lambda^{-1}v)$  of angle, delay, and Doppler-shift parameters, which appear in formula (2). The angle, delay, and Doppler-shift information can always easily be transformed back into the physical coordinates.

## Discretization

We switch to a discrete setting by considering sampled (at Nyquist rate) versions of the transmitted signals and, furthermore, assuming that the targets are located on a grid in the angle-delay-Doppler domain. In certain applications, this idealizing assumption might lead to so-called “gridding errors”. Thus, in practice, this would have to be resolved by an additional post-processing. Theoretical approaches for this problem can be found in [18, 9]. Recently, in [16], a new approach for exact reconstruction of time/frequency shifts (as considered in radar) from a continuous domain has been proposed.

Let  $\mathbf{s}_i \in \mathbb{C}^{N_t}$  be the discrete-time representation of the transmitted signal  $s_i = s_i(t)$  sampled at Nyquist-rate over the time interval  $[0, T)$ , i.e.,

$$\mathbf{s}_i^T = (s_i(0), s_i(\Delta_t), s_i(2\Delta_t), \dots, s_i((N_t - 1)\Delta_t))^T, \quad T = N_t \Delta_t, \quad \Delta_t = 1/2B.$$

Under the assumption that the signals  $s_i = s_i(t)$  are  $T$ -periodic and band-limited to  $(-B, B)$  the Nyquist-Shannon sampling theorem tells us that they are fully described by the discrete-time signals  $\mathbf{s}_i$ . In the following we will always assume that the parameters  $(\sin(\theta), 2r/c, 2\lambda^{-1}v)$  lie on an equidistant grid, i.e.,

$$(\sin(\theta), 2r/c, 2\lambda^{-1}v) = (\beta\Delta_\beta, \tau\Delta_\tau, f\Delta_f), \quad (3)$$

where  $\beta$ ,  $\tau$ , and  $f$  are integers. Throughout the rest of this paper we fix the stepsizes  $\Delta_\beta$ ,  $\Delta_\tau$ , and  $\Delta_f$  to

$$\Delta_\beta = \frac{2}{N_T N_R}, \quad \Delta_\tau = \frac{1}{2B}, \quad \Delta_f = \frac{1}{T}. \quad (4)$$

For a given Doppler-shift  $2\lambda^{-1}v = f\Delta_f \ll B$ , the received signal  $y_j = y_j(t)$  can again be assumed to be band-limited to  $(-B, B)$ . In this case the time-discrete version

$$\mathbf{y}_j^T = (y_j(0), y_j(\Delta_t), y_j(2\Delta_t), \dots, y_j((N_t - 1)\Delta_t))^T \in \mathbb{C}^{N_t}$$

of the approximate received signal  $y_j = y_j(t)$  in (2) at the  $j$ th receiver, caused by a target having parameters as in (3), is given by the vector

$$\mathbf{y}_j = e^{i2\pi \cdot c\lambda^{-1}\tau\Delta_\tau} e^{i2\pi \cdot d_R\beta\Delta_\beta(j-1)} \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T\beta\Delta_\beta(i-1)} \mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i \in \mathbb{C}^{N_t}, \quad (5)$$

where  $\mathbf{T}_\tau$  is a circulant shift and  $\mathbf{M}_f$  is a linear modulation, defined entrywise by

$$[\mathbf{T}_\tau \mathbf{s}]_k = [\mathbf{s}]_{k-\tau}, \quad [\mathbf{M}_f \mathbf{s}]_k = e^{i2\pi \cdot \frac{f}{N_t}(k-1)} [\mathbf{s}]_k, \quad (6)$$

where  $k - \tau$  stands for subtraction modulo  $N_t$ . Note that, due to periodicity of the operators  $\mathbf{T}_\tau$  and  $\mathbf{M}_f$  and the periodic influence of the parameter  $\beta$  in (5), the maximal achievable number of grid points as in (3) which can be differentiated by looking at the received signals  $\mathbf{y}_j$  is bounded by  $N_T N_R N_t^2$ . Hence we assume without loss of generality that

$$(\beta, \tau, f) \in \mathcal{G}, \quad \mathcal{G} := [N_T N_R] \times [N_t] \times [N_t], \quad (7)$$

where, for any  $L \in \mathbb{N}$ , we write  $[L]$  to denote the set  $\{1, 2, \dots, L\}$ . Throughout this paper we will refer to the triples of integers  $(\beta, \tau, f)$  as *parameters* of the considered targets keeping in mind that the actual *physical* parameters can be obtained via the relations (3), (4).

## The linear measurement model

If, more generally, several targets  $\Theta_k = (\beta_k, \tau_k, f_k)$ ,  $k = 1, 2, \dots, L$ , with corresponding reflectivities  $\rho_k \in \mathbb{C}$  are present, then the superposition of signals

$$\mathbf{y}_j = \sum_{k=1}^L \rho_k \cdot e^{i2\pi \cdot c\lambda^{-1}\tau_k\Delta_\tau} \left[ e^{i2\pi \cdot d_R\beta_k\Delta_\beta(j-1)} \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T\beta_k\Delta_\beta(i-1)} \mathbf{M}_{f_k} \mathbf{T}_{\tau_k} \mathbf{s}_i \right] \in \mathbb{C}^{N_t}$$

will be recorded at the  $j$ th receiver, see also (5). We define vectors  $\mathbf{A}_\Theta^j \in \mathbb{C}^{N_t}$  with

$$\mathbf{A}_\Theta^j = e^{i2\pi \cdot d_R \beta \Delta_\beta(j-1)} \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T \beta \Delta_\beta(i-1)} \mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i, \quad \Theta = (\beta, \tau, f) \in \mathcal{G}, \quad (8)$$

which allows us to write  $\mathbf{y}_j$  more conveniently as the sum  $\sum_{k=1}^L \rho_k e^{i2\pi \cdot c \lambda^{-1} \tau_k \Delta_\tau} \mathbf{A}_{\Theta_k}^j$ . A *target scene* (a set of targets being present in the angle-delay-Doppler domain) can now equivalently be considered as a vector  $\mathbf{x} = (x_\Theta, \Theta \in \mathcal{G}) \in \mathbb{C}^N$ ,  $N = N_T N_R N_t^2$ , being supported on the indices  $\Theta = \Theta_k$ ,  $k = 1, 2, \dots, L$ , which correspond to the parameters  $\Theta_k$  of the targets. Here, each nonzero entry is basically the reflectivity parameter of the corresponding target (multiplied with a complex unit), given by  $x_{\Theta_k} = \rho_k e^{i2\pi \cdot c \lambda^{-1} \tau_k \Delta_\tau}$ . The collection of all received signals  $(\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_{N_R}^T)$  can now conveniently be written as matrix-vector product  $\mathbf{A}\mathbf{x}$ , where the matrix  $\mathbf{A} \in \mathbb{C}^{N_R N_t \times N}$  contains the columns  $\mathbf{A}_\Theta$ ,  $\Theta \in \mathcal{G}$ , with

$$\mathbf{A}_\Theta = ((\mathbf{A}_\Theta^1})^T, (\mathbf{A}_\Theta^2})^T, \dots, (\mathbf{A}_\Theta^{N_R})^T)^T \in \mathbb{C}^{N_R N_t}. \quad (9)$$

Considering noise in the channel leads to the measurement model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \in \mathbb{C}^{N_R N_t}, \quad (10)$$

where  $\mathbf{n}$  represents a noise vector. Since  $\mathbf{x}$  is an  $N = N_R N_T N_t^2$ -dimensional vector and the acquired information in the measured vector  $\mathbf{y}$  is  $m = N_R N_t$ -dimensional, we end up with a highly under-determined system (10) from which the targets' parameters (the support set of the vector  $\mathbf{x}$ ) have to be reconstructed.

## 1.2 Recovery via $\ell_1$ -minimization and the restricted isometry property

Radar scenes are typically sparse in the target domain meaning  $s = |\text{supp}(\mathbf{x})| \ll N$ . One can use this additional assumption for the reconstruction of  $\mathbf{x}$ . At this point ideas from compressed sensing enter the field. A well established — and computationally practicable — approach for reconstructing a sparse vector  $\mathbf{x}$  from incomplete measurements  $\mathbf{y}$  is *basis pursuit denoising* which aims at reconstructing  $\mathbf{x}$  by solving the convex optimization problem

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \varrho, \quad (11)$$

where  $\varrho$  is an upper estimate of the noise level. Here the  $\|\cdot\|_1$ -norm constitutes a convex relaxation of the  $\|\cdot\|_0$ -norm counting the number of nonzero entries, which one aims to minimize in the first place. Indeed, under suitable assumptions on the measurement matrix  $\mathbf{A}$ , each solution  $\mathbf{x}^\#$  of (11) is close to the unknown  $s$ -sparse vector  $\mathbf{x}$ , i.e.,

$$\|\mathbf{x}^\# - \mathbf{x}\|_1 \leq C \inf_{s\text{-sparse } \mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|_1 + D\varrho\sqrt{s}, \quad (12)$$

$$\|\mathbf{x}^\# - \mathbf{x}\|_2 \leq C \frac{\inf_{s\text{-sparse } \mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|_1}{\sqrt{s}} + D\varrho, \quad (13)$$

where  $C, D > 0$  are numerical constants.

A popular condition guaranteeing recovery via basis pursuit denoising is the *restricted isometry property (RIP)*. The RIP is said to be fulfilled if  $\mathbf{A}$  has a small restricted isometry constant which is the smallest number  $\delta_s$  such that for all  $s$ -sparse vectors  $\mathbf{x}$  it holds

$$(1 - \delta_s) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_s) \|\mathbf{x}\|_2^2. \quad (14)$$

A sufficiently small restricted isometry constant implies that the basis pursuit denoising approach (11) yields approximations to any  $s$ -sparse vector  $\mathbf{x}$  from incomplete measurements  $\mathbf{y}$ . The following theorem is well-known [4], see for instance also [7, 5, 14], where constants have been subsequently been improved.

**Theorem 1.** *Suppose the restricted isometry constant  $\delta_{2s}$  of the matrix  $\mathbf{A}$  satisfies  $\delta_{2s} < 1/\sqrt{2}$ . Then, for any  $\mathbf{x} \in \mathbb{C}^N$  and  $\mathbf{y} \in \mathbb{C}^m$  with  $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \varrho$ , any solution  $\mathbf{x}^\#$  of (11) fulfills (12), (13) where the constants  $C, D > 0$  only depend on  $\delta_{2s}$ .*

For certain random matrices, e.g., those with independent standard Gaussian entries, the following choice of the number of measurements  $m$  depending on the dimension  $N$  and the sparsity level  $s$  is sufficient (and necessary, see below) for a small restricted isometry constant with high probability [14]:

$$m \gtrsim s \log(eN/s). \quad (15)$$

Apart from Gaussian measurements this dependence also occurs for many other (even structured) random matrices — up to possibly additional logarithmic factors.

### 1.3 An optimal RIP result

A main goal of this paper consists in proving the following result on the number of required samples  $N_t$  guaranteeing a small restricted isometry constant of the scaled MIMO radar measurement matrix (see (8), (9) for the definition)

$$\tilde{\mathbf{A}} = \frac{1}{\sqrt{N_T N_R N_t}} \mathbf{A} \in \mathbb{C}^{N_R N_t \times N},$$

and, moreover, in showing that this result is optimal in a certain sense. We assume that the transmit pulse vectors  $\mathbf{s}_i$  are independent standard complex Gaussian random vectors (see Appendix D for some information about complex Gaussian random variables and vectors).

**Theorem 2.** *Let the signals  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_T}$  generating the matrix  $\mathbf{A}$  via (8) and (9) be independent standard complex Gaussian random vectors. If, for  $\delta, \varepsilon \in (0, 1)$ ,*

$$N_t \gtrsim \frac{s}{\delta^2} \max\{\log^2(eN) \log^2(s), \log(1/\varepsilon)\}, \quad (16)$$

*then the restricted isometry constant of the rescaled matrix  $\tilde{\mathbf{A}}$  satisfies  $\delta_s(\tilde{\mathbf{A}}) < \delta$  with probability at least  $1 - \varepsilon$ .*

**Remark 3.** *The result as well as its proof extends to signals  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_T}$  being independent Rademacher vectors (random vectors with independent entries taking the value  $+1$  or  $-1$  with equal probability) or independent Steinhaus vectors (random vectors with independent entries that are uniformly distributed on the complex torus). These generators may be of advantage for real radar systems where one prefers constant magnitude signals for optimal energy consumption and in order to avoid that amplifiers possibly run out of their linear regime.*

Recall that the matrix  $\tilde{\mathbf{A}}$  has  $m = N_R N_t$  rows, i.e.,  $\tilde{\mathbf{A}}$  represents  $N_R N_t$  measurements. In this sense the theorem may seem suboptimal considering that one typically would expect a scaling as in (15). However, using general theoretical results on lower bounds for the necessary number of measurements, we argue below that our RIP result above is in fact optimal (up to possibly logarithmic factors).

#### Instance optimality

The concept of *instance optimality* [10, 14] allows to derive general lower bounds on the number of measurements  $m$  required for stable sparse reconstruction via any algorithm and any possible measurement matrix in  $\mathbb{C}^{m \times N}$ . To be precise, we say that a pair  $(\mathbf{A}, \Delta)$ , where  $\Delta: \mathbb{C}^m \rightarrow \mathbb{C}^N$  is a reconstruction map and  $\mathbf{A} \in \mathbb{C}^{m \times N}$ , is  $\ell_1$ -instance optimal of order  $s$  with constant  $C > 0$  if for all  $\mathbf{x} \in \mathbb{C}^N$  it holds

$$\|\mathbf{x} - \Delta(\mathbf{A}\mathbf{x})\|_1 \leq C \inf_{s\text{-sparse } \mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|_1.$$

The following result states the announced lower bound.

**Theorem 4** ([14, Thm. 11.6]). *If a pair of measurement matrix  $\mathbf{A} \in \mathbb{C}^{m \times N}$  and reconstruction map  $\Delta: \mathbb{C}^m \rightarrow \mathbb{C}^N$  is  $\ell_1$ -instance optimal of order  $s$  with constant  $C$ , then*

$$m \geq C' s \log(eN/s) \quad (17)$$

*for some constant  $C'$  depending only on  $C$ .*

In case of the MIMO radar measurement matrix  $\mathbf{A}$ , this result can be applied to a certain submatrix of  $\mathbf{A}$ . To this end, we define  $\mathcal{G}_{\beta^*} := \{(\beta, \tau, f) \in \mathcal{G} : \beta = \beta^*\}$ , for an arbitrary but fixed angle parameter  $\beta^*$ . For a target scene  $\mathbf{x}$  with support in  $\mathcal{G}_{\beta^*}$  one finds (recalling (8)) that the product  $\mathbf{A}\mathbf{x}$  can be expressed as outer product  $\mathbf{w} \otimes \mathbf{B}\mathbf{x}_{\mathcal{G}_{\beta^*}}$ , where  $\mathbf{B}$  is an  $N_t \times N_t^2$  matrix with columns

$$\mathbf{B}_{(\beta^*, \tau, f)} = \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T \beta^* \Delta_\beta(i-1)} \mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i, \quad (\beta^*, \tau, f) \in \mathcal{G}_{\beta^*},$$

and  $\mathbf{w}$  is a  $N_R$ -dimensional vector, entrywise defined as

$$[\mathbf{w}]_j = e^{i2\pi \cdot d_R \beta^* \Delta_\beta(j-1)}, \quad j \in [N_R],$$

and the vector  $\mathbf{x}_{\mathcal{G}_{\beta^*}}$  is obtained from  $\mathbf{x}$  by deleting all entries which do not belong to the set  $\mathcal{G}_{\beta^*}$ . This implies that  $\|\tilde{\mathbf{A}}\mathbf{x}\|_2^2 = \|\tilde{\mathbf{B}}\mathbf{x}_{\mathcal{G}_{\beta^*}}\|_2^2$ , where  $\tilde{\mathbf{B}}$  stands for the scaled matrix  $\frac{1}{\sqrt{N_T N_t}} \mathbf{B}$ . Consequently, the restricted isometry constant  $\delta_{2s}(\tilde{\mathbf{B}})$  is bounded by  $\delta_{2s}(\tilde{\mathbf{A}})$ . We conclude that  $\delta_{2s}(\tilde{\mathbf{A}}) < 1/\sqrt{2}$  implies  $\delta_{2s}(\tilde{\mathbf{B}}) < 1/\sqrt{2}$  so that by Theorem 1 the matrix  $\tilde{\mathbf{B}}$  in combination with the reconstruction map corresponding to basis pursuit (11) with  $\rho = 0$  is  $\ell_1$ -instance optimal. Since  $\tilde{\mathbf{B}}$  has  $N_t$  rows and  $N_t^2$  columns, (17) reads as  $N_t \gtrsim s \log(eN_t^2/s)$ . Therefore, we can formulate the following result. (Essentially, this result could have similarly been derived from Corollary 10.8 in [14].)

**Theorem 5.** *If the restricted isometry constant of the scaled MIMO radar measurement matrix  $\tilde{\mathbf{A}} \in \mathbb{C}^{N_R N_t \times N}$  considered in Theorem 2 satisfies  $\delta_{2s} < 1/\sqrt{2}$ , then it holds*

$$N_t \gtrsim s \log(eN_t^2/s). \quad (18)$$

This shows that — up to additional logarithmic factors — Theorem 2 on the restricted isometry constant for the MIMO radar measurement matrix is indeed optimal. Analogously one can also argue that  $\ell_1$ -instance optimal uniform recovery in general — which is not necessarily based on the RIP — is not possible with less measurements than in (16). Indeed, by assuming that there exists a corresponding reconstruction map for the matrix  $\tilde{\mathbf{A}}$  one can use this map for constructing an  $\ell_1$ -instance optimal reconstruction map for any submatrix in  $\tilde{\mathbf{B}}$  as defined above, which again yields (18).

## 1.4 Nonuniform recovery for random support sets

As stated by Theorem 2, for a *fixed* draw  $\mathbf{A}$  of the measurement matrix the RIP is fulfilled with high probability and, hence (due to Theorem 1), *all* sparse vectors  $\mathbf{x}$  can be reconstructed by considering measurements  $\mathbf{A}\mathbf{x}$ . Since in this scenario  $\mathbf{A}$  is fixed and reconstruction is possible for all sparse vectors  $\mathbf{x}$ , we speak of a *uniform* recovery result. A *nonuniform* recovery result, on the other hand, only states that a given sparse vector  $\mathbf{x}$  can be reconstructed from measurements  $\mathbf{A}\mathbf{x}$  with high probability on the draw of the matrix  $\mathbf{A}$ , i.e., no assertion on the reconstruction of all sparse vectors using a single draw of the matrix  $\mathbf{A}$  is made.

The results on the MIMO radar measurements by Strohmer & Friedlander [27] imply that, at least on average, *nonuniform* reconstruction (for random support sets) succeeds with less measurements. In [27] they derive results for nonuniform reconstruction via the *debiased LASSO* reconstruction scheme. In case of the *standard* LASSO, an approximation  $\mathbf{x}^\#$  of  $\mathbf{x}$  is obtained by solving the minimization problem

$$\min_{\mathbf{z}} \frac{1}{2} \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}\|_1, \quad (19)$$

where  $\lambda > 0$  has to be chosen appropriately (in accordance with the noise level). In case that the support  $S = \text{supp}(\mathbf{x})$  is recovered correctly, i.e.,  $\text{supp}(\mathbf{x}^\#) = S$ , one might add a further step consisting of solving the reduced least squares problem

$$\min_{\mathbf{z}} \|\mathbf{A}_S \mathbf{z} - \mathbf{y}\|_2^2, \quad (20)$$

where  $\mathbf{A}_S$  contains only the columns of  $\mathbf{A}$  corresponding to the support set  $S$  and the minimization is over all  $\mathbf{z} \in \mathbb{C}^{|S|}$ . This second “debiasing” step leads in general to an improvement of the reconstructed



coefficients. The LASSO approach is closely related to basis pursuit denoising (11) but due to the unconstrained optimization often favorable in practice [14]. Strohmer & Friedlander assume a random target model, assuming that for a fixed sparsity parameter every support set occurs with the same probability and, moreover, the phases of the nonzero entries of the targets are independent and uniformly distributed in  $[0, 2\pi)$ . This model has been studied in [6] in a general context and is called the *generic s-sparse target model* therein; see also [14, Chap. 14]. Under additional technical assumptions, it has been shown in [27]<sup>1</sup> that if  $\mathbf{x}$  is an  $s$ -sparse target scene, then

$$N_R N_t \gtrsim s \log(N), \quad (21)$$

measurements are sufficient to guarantee that, with high probability, each solution  $\mathbf{x}^\#$  to the LASSO has in particular the exact same support set as the target vector  $\mathbf{x}$ , thus perfectly recovering the parameters of the targets.

## 1.5 Nonuniform recovery for deterministic support sets

The results by Strohmer & Friedlander imply that there must exist support sets which can be reconstructed using less measurements (than in our RIP result, see Theorem 2). As it turns out, certain angle classes are not correlated among each other and, hence, support sets where the mass is uniformly distributed among these angle classes are favorable for reconstruction.

### Balanced support sets

Recall the definition of the columns  $\mathbf{A}_\Theta$ ,  $\Theta \in \mathcal{G}$ , of the MIMO radar measurement matrix  $\mathbf{A}$  from (8), (9). For the inner product of two such columns  $\mathbf{A}_\Theta, \mathbf{A}_{\Theta'}$  it holds

$$\langle \mathbf{A}_\Theta, \mathbf{A}_{\Theta'} \rangle = \sum_{k=1}^{N_R} e^{i2\pi \cdot d_R(\beta' - \beta)\Delta_\beta(k-1)} \left\langle \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T \beta \Delta_\beta(i-1)} \mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i, \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T \beta' \Delta_\beta(i-1)} \mathbf{M}_{f'} \mathbf{T}_{\tau'} \mathbf{s}_i \right\rangle. \quad (22)$$

Due to the fact that  $d_R = N_T/2$  and  $\Delta_\beta = 2/N_T N_R$ , the first factor can be calculated as

$$\sum_{k=1}^{N_R} e^{i2\pi \cdot d_R(\beta' - \beta)\Delta_\beta(k-1)} = \delta_{\beta, \beta'}^{\sim N_R} N_R, \quad \text{where} \quad \delta_{\beta, \beta'}^{\sim N_R} := \begin{cases} 1, & \text{if } \beta' - \beta \in N_R \mathbb{Z}, \\ 0, & \text{else.} \end{cases}$$

This means that for two columns  $\mathbf{A}_{(\beta, \tau, f)}$ ,  $\mathbf{A}_{(\beta', \tau', f')}$  to be correlated it is necessary that  $\beta' - \beta \in N_R \mathbb{Z}$ . The latter condition induces an equivalence relation on the set of possible angle parameters. We say that two angle parameters  $\beta, \beta'$  are equivalent if and only if  $\beta' - \beta \in N_R \mathbb{Z}$ , i.e.,

$$\beta' \sim \beta \iff \beta' - \beta \in N_R \mathbb{Z},$$

which, recalling from (7) that  $\beta$  is taken from  $\{1, 2, \dots, N_T N_R\}$ , leaves us with  $N_R$  angle classes  $[\beta]$ , each containing  $N_T$  different angle parameters. The crucial fact is that, since the columns from different angle classes are uncorrelated, the matrices  $\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S$ ,  $S \subset \mathcal{G}$ , are block diagonal. It holds in particular,

$$\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} = \max_{[\beta]} \|\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}} - \mathbf{Id}\|_{2 \rightarrow 2}, \quad (23)$$

where the subsets  $S_{[\beta]} \subseteq S$  are defined as

$$S_{[\beta]} := \{\Theta' = (\beta', \tau', f') \in S : \beta' \sim \beta\}. \quad (24)$$

Considering (23) one may expect that a uniform distribution of any given support set  $S$  over all angle classes would make it most likely that the matrix  $\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S$  is well conditioned (since in this case the maximal dimension of the submatrices  $\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}}$  becomes minimal). This motivates to introduce a parameter  $\eta$  measuring how evenly distributed over the angle classes such a support set is.

<sup>1</sup>Note that there is a typo in [27, Thm. 5] where an additional factor  $N_t$  on the left-hand side of (21) shows up.



**Definition 6.** The balancedness parameter of a support set  $S \subset \mathcal{G}$  is the smallest number  $\eta$  such for all angle classes  $[\beta]$  it holds

$$|S_{[\beta]}| \leq \eta \frac{|S|}{N_R}.$$

We say that  $S$  is  $\eta$ -balanced, if the balancedness parameter of  $S$  equals  $\eta$ .

According to this definition the balancedness parameter can be expressed as

$$\eta = \max_{[\beta]} \frac{N_R |S_{[\beta]}|}{|S|}.$$

Clearly, the balancedness parameter takes values in  $[1, N_R]$ , where  $\eta = 1$  means that the  $|S|$  targets are perfectly balanced with respect to the  $N_R$  angle classes (with  $|S_{[\beta]}| \sim |S|/N_R$ ), and, on the other hand,  $\eta = N_R$  corresponds to the case where all targets are located in one particular angle class.

### Nonuniform recovery results

Next we present our nonuniform recovery results for  $s$ -sparse target scenes with  $\eta$ -balanced support sets (see Definition 6 above) and, thereby, reveal an essentially linear dependence of the number of measurements  $m = N_R N_t$  required for reconstruction on the number of targets  $s$ . Note, how the parameter  $\eta$  enters the corresponding condition (25) below. In the worst possible case where  $\eta = N_R$  we obtain essentially the same scaling as for the RIP estimate in (16), while for the best possible case  $\eta = 1$  we obtain the announced linear scaling.

For the following we recall that a Steinhaus sequence is a sequence of independent random variables that are uniformly distributed on the complex unit circle  $\{z \in \mathbb{C} : |z| = 1\}$ . Due to the fact that the grid  $\mathcal{G}$  (see (7)) is  $N = N_T N_R N_t^2$ -dimensional, a target scene  $\mathbf{x}$  can be regarded as a vector in  $\mathbb{C}^N$ .

**Theorem 7.** Let  $\mathbf{x} \in \mathbb{C}^N$  be an  $s$ -sparse target scene with  $\eta$ -balanced support set and assume that measurements  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \in \mathbb{C}^{N_R N_t}$  are given, where the signals  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_T}$  generating the columns of  $\mathbf{A}$  (according to (8), (9)) are independent standard complex Gaussian random vectors, and the noise vector  $\mathbf{n}$  is a mean-zero complex Gaussian random vector with variance  $\sigma^2$ . If

$$N_R N_t \gtrsim \eta s \log^3(N/\varepsilon), \quad (25)$$

and the signs of the nonzero entries of  $\mathbf{x}$  form a Steinhaus sequence while the magnitudes satisfy

$$\min_{\Theta \in \text{supp}(\mathbf{x})} |\mathbf{x}_\Theta| > \frac{8\sigma}{\sqrt{N_T N_R N_t}} \sqrt{2 \log(N)}, \quad (26)$$

then, with probability at least  $1 - 7 \max\{\varepsilon, N^{-2}\}$ ,

(a) the solution  $\mathbf{x}^\#$  to the LASSO (19) with  $\lambda = 2\sigma \sqrt{2 N_T N_R N_t \log(N)}$  satisfies

$$\text{supp}(\mathbf{x}^\#) = \text{supp}(\mathbf{x}),$$

(b) the solution  $\mathbf{z}^\#$  to the least squares problem (20) with  $S = \text{supp}(\mathbf{x}^\#)$  satisfies

$$\|\mathbf{z}^\# - \mathbf{x}_S\|_2 \leq \frac{2\sigma\sqrt{s}}{\sqrt{N_T N_R N_t}} \sqrt{2 \log(N)}. \quad (27)$$

Note that, in view of the scaling of  $s$  in (25), the approximation estimate (27) actually implies that  $\|\mathbf{z}^\# - \mathbf{x}_S\|_2 \lesssim \sigma/\sqrt{N_T}$ , where  $N_T$  is the number of transmitters. Furthermore, we would like to point out that, for technical reasons, we did not consider deterministic signs of the nonzero entries of  $\mathbf{x}$ . Nevertheless, we do believe that with similar ideas as in [19] such a further generalization would be possible (although technical).

The next theorem, our second main result on nonuniform recovery, makes an assertion about reconstruction via basis pursuit denoising. Unlike the previous Theorem 7, the following result deals with target scenes which are not exactly sparse but rather can be approximated well by sparse vectors. Again, the balancedness parameter  $\eta$  of the considered support sets is essential.

**Theorem 8.** Let  $\mathbf{x} \in \mathbb{C}^N$  be a target scene, let  $S \subset \mathcal{G}$  be an index set corresponding to  $s$  largest (by magnitude) entries in  $\mathbf{x}$ , and let  $S$  be  $\eta$ -balanced. Further, assume that the signs of the coefficients  $\mathbf{x}_S$  form a Steinhaus sequence. Assume measurements  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \in \mathbb{C}^{N_R N_t}$  are given, where the signals  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_T}$  generating the columns of the measurement matrix  $\mathbf{A}$  (according to (8), (9)) are independent standard Gaussian random vectors, and the noise vector  $\mathbf{n}$  is known to satisfy  $\|\mathbf{n}\|_2 \leq \varrho$ . If

$$N_R N_t \gtrsim \eta s \log^3(N/\varepsilon), \quad (28)$$

then, with probability at least  $1 - \varepsilon$ , the solution  $\mathbf{x}^\#$  to the basis pursuit denoising program (11) satisfies

$$\|\mathbf{x}^\# - \mathbf{x}\|_2 \leq C_1 \inf_{s\text{-sparse } \mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|_1 + C_2 \frac{\varrho \sqrt{s}}{\sqrt{N_T N_R N_t}},$$

where  $C_1$  and  $C_2$  are numerical constants.

**Remark 9.** Like Theorem 2 concerning the RIP, Theorem 7 and 8 extend to signals  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_T}$  being independent Rademacher and Steinhaus sequences, compare also with Remark 3.

## 1.6 The Doppler-free scenario

In [27] also a Doppler-free scenario (for the case of slowly moving or stationary targets) has been analyzed. We want to point out that this case is also covered by our analysis. The measurement matrix  $\mathbf{A}'$  corresponding to the Doppler-free case can be obtained from  $\mathbf{A}$  by deleting all columns  $\mathbf{A}_{(\beta, \tau, f)}$  with nonzero Doppler shifts  $f$ . Therefore, in view of (14), our RIP result (see Theorem 2) holds also automatically true for the Doppler-free case — where one may even replace the factor  $\log(N) = \log(N_T N_R N_t^2)$  by  $\log(N_T N_R N_t)$ . The proof of our nonuniform recovery results, stated as Theorems 7 and 8, is based on Proposition 13 providing estimates for the singular values of the matrices  $\mathbf{A}_S$  where  $S$  stands for an arbitrary (but balanced) support set. These estimates, of course, hold in particular true when considering only matrices  $\mathbf{A}_S$  with the property that  $S$  does not contain any index  $(\beta, \tau, f)$  with  $f \neq 0$  and is balanced. The reader might assure himself that all remaining arguments in Section 3 hold equally true for the Doppler-free case and, thus, analog results on nonuniform recovery can be proven.

## Outline

In Section 2 we prove the RIP result (Theorem 2) by reformulating the restricted isometry constant as the supremum of a chaos process and then applying a general result on such processes shown in [20, 11]. Section 3 is devoted to the proofs of our nonuniform recovery result for balanced support sets (Theorems 7 and 8). The proof of Theorem 7 is based on a general reconstruction result for the LASSO approach taken from [6] which we introduce in Section 3.1. The Sections 3.1.1–3.1.3 are devoted to the verification of the assumption needed for applying the reconstruction result. To this end a probabilistic analysis of the extremal singular values of the matrices  $\tilde{\mathbf{A}}_S$ , which we state as Proposition 13, is essential. Since the proof of Proposition 13 is rather extensive it appears in an extra section, namely Section 3.3. In Section 3.2 we prove Theorem 8 by applying a more general recovery result for the basis pursuit denoising approach, taken from [14]. Also in this case the verification of the assumptions depends crucially on the analysis of the singular values of  $\tilde{\mathbf{A}}_S$  provided by Proposition 13. Finally, Section 4 is devoted to some numerical experiments supporting our theoretical results on the influence of the balancedness parameter  $\eta$  (of the considered support sets) on the recovery performance.

In the Appendix we list basic calculations and technical proofs which for the sake of a clearer presentation do not appear in the main part. Furthermore, some tools from probability theory and a short introduction of standard complex Gaussian random variables and vectors can be found here.

## Notation

For a complex number  $z$  we write  $\bar{z}$  for the complex conjugate and, for nonzero  $z$ , we set  $\text{sgn}(z) := z/|z|$ . Given a vector  $\mathbf{x} = (x_1, x_2, \dots)^T$  with complex entries, we write  $\|\mathbf{x}\|_p = (\sum_k |x_k|^p)^{1/p}$ ,  $1 \leq p < \infty$ , to denote the usual  $\ell_p$ -norm of  $\mathbf{x}$ , and  $\|\mathbf{x}\|_\infty = \max_k |x_k|$ . Moreover,  $\text{supp}(\mathbf{x}) = \{k : x_k \neq 0\}$  denotes the support set of  $\mathbf{x}$ . Occasionally we also write  $[\mathbf{x}]_k$  to denote the  $k$ th entry of the vector  $\mathbf{x}$ . We write

**Id** for the identity matrix (with appropriate dimensions becoming clear from the context). Given a complex valued matrix  $\mathbf{A}$ , we write  $[\mathbf{A}]_{k,\ell}$  to denote the  $k$ th entry of the  $\ell$ th row of  $\mathbf{A}$ . The Hermitian transpose will be denoted by  $\mathbf{A}^*$ . The spectral norm is denoted by  $\|\mathbf{A}\|_{2 \rightarrow 2} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2$ . The Frobenius norm of a matrix  $\mathbf{A}$  is given by  $\|\mathbf{A}\|_F = \sqrt{\sum_{k,\ell} [\mathbf{A}]_{k,\ell}^2}$ . Given the sequence of singular values  $\sigma(\mathbf{A}) = (\sigma_1(\mathbf{A}), \sigma_2(\mathbf{A}), \dots)$  of a matrix  $\mathbf{A}$ , the Schatten  $p$ -norm, for  $1 \leq p \leq \infty$ , is given by  $\|\mathbf{A}\|_{S_p} := \|\sigma(\mathbf{A})\|_p$ . Note the special cases  $\|\mathbf{A}\|_{S_2} = \|\mathbf{A}\|_F$  and  $\|\mathbf{A}\|_{S_\infty} = \|\mathbf{A}\|_{2 \rightarrow 2}$ . For a set  $\mathcal{A}$  of matrices, the parameters  $d_{2 \rightarrow 2}(\mathcal{A}) = \sup_{\mathbf{A} \in \mathcal{A}} \|\mathbf{A}\|_{2 \rightarrow 2}$ ,  $d_F(\mathcal{A}) = \sup_{\mathbf{A} \in \mathcal{A}} \|\mathbf{A}\|_F$  denote the diameters of  $\mathcal{A}$  with respect to the spectral norm and the Frobenius norm, respectively. Given a number  $L \in \mathbb{N}$ , we write  $[L]$  to denote the set  $\{1, 2, \dots, L\}$ . We will also make use of the symbol  $\delta_{k,\ell}$  or, more generally,  $\delta_{k,\ell}^{\sim L}$ , for  $k, \ell \in \mathbb{Z}$  and  $L \in \mathbb{N}$ . Here  $\delta_{k,\ell}$  is the usual Kronecker delta which is equal to 1 if  $k = \ell$  and 0, otherwise. The symbol  $\delta_{k,\ell}^{\sim L}$ , on the other hand, represents equality up to multiples of  $L$ , i.e.,

$$\delta_{k,\ell}^{\sim L} = \begin{cases} 1 & \text{if } k - \ell \in L\mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases} \quad (29)$$

We write  $A \lesssim B$  if there is a constant  $c_1$  with  $A \leq c_1 B$ , where  $A, B$  may depend on further parameters.

## 2 Uniform recovery via the RIP

In this section we prove Theorem 2 concerning the RIP of the MIMO radar measurement matrix. In the following,  $\mathbf{s}$  denotes the vector containing all signals  $\mathbf{s}_i$ , i.e.,

$$\mathbf{s} = (\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_{N_T}^T)^T \in \mathbb{C}^{N_T N_t}.$$

### 2.1 Reformulation as a chaos process

Our proof is based on [20] where bounds for certain chaos processes have been developed. The key is to reformulate the radar measurements  $\mathbf{x} \mapsto \tilde{\mathbf{A}}\mathbf{x}$  as a mapping,  $\mathbf{x} \mapsto \tilde{\mathbf{V}}_{\mathbf{x}}\mathbf{s}$ , where the matrix  $\tilde{\mathbf{V}}_{\mathbf{x}}$ , depending on  $\mathbf{x}$ , is given by

$$\tilde{\mathbf{V}}_{\mathbf{x}} := \frac{1}{\sqrt{N_T N_R N_t}} \sum_{\theta \in \mathcal{G}} x_\theta \mathbf{X}_\theta, \quad (30)$$

and where  $\mathbf{X}_\theta$  is a  $N_R \times N_T$  block matrix consisting of the  $N_t \times N_t$  blocks  $\mathbf{X}_\theta^{[i,j]}$  with

$$\mathbf{X}_\theta^{[i,j]} := e^{i2\pi \cdot d_R \beta \Delta_\beta(i-1)} e^{i2\pi \cdot d_T \beta \Delta_\beta(j-1)} \mathbf{M}_f \mathbf{T}_\tau. \quad (31)$$

Due to the fact that the signal vector  $\mathbf{s}$  is isotropic, we have for each  $\mathbf{x}$  that  $\mathbb{E} \|\tilde{\mathbf{V}}_{\mathbf{x}}\mathbf{s}\|_2^2 = \|\tilde{\mathbf{V}}_{\mathbf{x}}\|_F^2$ , and, due to the fact that the matrices  $\frac{1}{\sqrt{N_T N_R N_t}} \mathbf{X}_\theta$ ,  $\theta \in \mathcal{G}$ , are orthogonal (see Appendix A),  $\|\tilde{\mathbf{V}}_{\mathbf{x}}\|_F^2 = \|\mathbf{x}\|_2^2$ . This enables us to express the restricted isometry constant  $\delta_s$  as supremum of a chaos process of the form

$$\delta_s = \sup_{\substack{\mathbf{s} \text{ s-sparse} \\ \|\mathbf{s}\|_2 \leq 1}} \left| \|\tilde{\mathbf{A}}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| = \sup_{\tilde{\mathbf{V}} \in \mathcal{A}} \left| \|\tilde{\mathbf{V}}\mathbf{s}\|_2^2 - \mathbb{E} \|\tilde{\mathbf{V}}\mathbf{s}\|_2^2 \right|, \quad (32)$$

where  $\mathcal{A}$  denotes the set of matrices given by

$$\mathcal{A} := \{ \tilde{\mathbf{V}}_{\mathbf{x}} : \mathbf{x} \text{ is s-sparse, } \|\mathbf{x}\|_2 \leq 1 \}. \quad (33)$$

Dudley's entropy integral

$$\mathcal{D} := \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)} du \quad (34)$$

for the set  $\mathcal{A}$ , where  $\mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)$  denotes the covering numbers, i.e., the minimal number of  $\|\cdot\|_{2 \rightarrow 2}$ -balls of radius  $u$  required to cover  $\mathcal{A}$ , provides sufficient information about the complexity of the set

$\mathcal{A}$  for us. Estimates for the parameter  $\mathcal{D}$  will in fact lead to good bounds for the restricted isometry constant in (32). The following theorem — which is a direct implication of [20, Thm. 3.1] and a slight improvement due to Dirksen [11] — is our crucial tool. Recall that a random vector  $\mathbf{X}$  is  $L$ -subgaussian if it is isotropic and  $\mathbb{P}(|\langle \mathbf{X}, \boldsymbol{\xi} \rangle| \geq t) \leq 2 \exp(-t^2/2L^2)$  for every  $\boldsymbol{\xi}$  with  $\|\boldsymbol{\xi}\|_2 = 1$  and any  $t > 0$ .

**Theorem 10.** *Let  $\mathcal{A}$  be a symmetric set of matrices,  $-\mathcal{A} = \mathcal{A}$ , and let  $\boldsymbol{\xi}$  be a random vector whose entries are independent, mean-zero, variance 1, and  $L$ -subgaussian random variables. Set*

$$E = \mathcal{D}(\mathcal{D} + d_F(\mathcal{A})),$$

$$V = d_{2 \rightarrow 2}(\mathcal{A})d_F(\mathcal{A}),$$

$$U = d_{2 \rightarrow 2}^2(\mathcal{A}).$$

Then, for  $t > 0$ ,

$$\mathbb{P}\left(\sup_{\mathbf{A} \in \mathcal{A}} \left| \|\mathbf{A}\boldsymbol{\xi}\|_2^2 - \mathbb{E}\|\mathbf{A}\boldsymbol{\xi}\|_2^2 \right| \geq c_1 E + t\right) \leq 2 \exp(-c_2 \min\{t^2/V^2, t/U\}),$$

where the constants  $c_1, c_2$  depend only on  $L$ .

Note that the original result [20, Thm. 3.1] is formulated in terms of Talagrand's chaining functional  $\gamma_2$ , whereas Theorem 10 from above is formulated in terms of Dudley's entropy integral  $\mathcal{D}$ , a common upper estimate for the  $\gamma_2$ -functional, which is sufficient for our purposes and more convenient. For more details on  $\gamma_2$ -functionals we refer to [28, 29]. In the original version of Theorem 10 in [20] the quantity  $V$  was defined as  $V = d_{2 \rightarrow 2}(\mathcal{A})(\mathcal{D} + d_F(\mathcal{A}))$ . Dirksen [11] achieved a slight improvement by showing that the summand  $\mathcal{D}$  can be omitted.

## 2.2 Proof of Theorem 2

In order to apply the above theorem we have to bound  $d_F(\mathcal{A})$ ,  $d_{2 \rightarrow 2}(\mathcal{A})$ , and  $\mathcal{D}$ . Estimates for the quantities  $d_{2 \rightarrow 2}(\mathcal{A})$  and  $d_F(\mathcal{A})$  can be derived in a straightforward manner. We have already seen that  $\|\tilde{\mathbf{V}}_{\mathbf{x}}\|_F = \|\mathbf{x}\|_2$  and, thus,  $d_F(\mathcal{A}) = 1$ . In order to provide a bound for  $d_{2 \rightarrow 2}(\mathcal{A})$ , we estimate the norms  $\|\boldsymbol{\chi}_{\Theta}\|_{2 \rightarrow 2}$  of the matrices from the definition in (30). The  $(j', j)$ th  $N_t \times N_t$  block of the product  $\boldsymbol{\chi}_{\Theta}^* \boldsymbol{\chi}_{\Theta}$  can be calculated (cf. Appendix A, equation (74)) as

$$[\boldsymbol{\chi}_{\Theta}^* \boldsymbol{\chi}_{\Theta}]^{[j', j]} = N_R e^{-i2\pi \cdot d_T \beta \Delta_{\beta}(j' - 1)} e^{i2\pi \cdot d_T \beta \Delta_{\beta}(j - 1)} \mathbf{Id}.$$

Thus, by defining a vector  $\mathbf{v} \in \mathbb{C}^{N_T}$  entrywise as

$$[\mathbf{v}]_j = e^{i2\pi \cdot d_T \beta \Delta_{\beta}(j - 1)}, \quad j \in [N_T],$$

the corresponding operator norm can be calculated as

$$\|\boldsymbol{\chi}_{\Theta}\|_{2 \rightarrow 2}^2 = \|\boldsymbol{\chi}_{\Theta}^* \boldsymbol{\chi}_{\Theta}\|_{2 \rightarrow 2} = N_R \|\mathbf{v} \mathbf{v}^*\|_{2 \rightarrow 2} = N_R \|\mathbf{v}\|_2^2 = N_R N_T.$$

Since for each  $s$ -sparse vector  $\mathbf{x}$  we have  $\|\mathbf{x}\|_1 \leq \sqrt{s} \|\mathbf{x}\|_2$ , we obtain

$$\|\tilde{\mathbf{V}}_{\mathbf{x}}\|_{2 \rightarrow 2} \leq \frac{1}{\sqrt{N_T N_R N_t}} \sum_{\Theta \in \mathcal{G}} |x_{\Theta}| \|\boldsymbol{\chi}_{\Theta}\|_{2 \rightarrow 2} \leq \frac{1}{\sqrt{N_t}} \|\mathbf{x}\|_1 \leq \sqrt{\frac{s}{N_t}} \|\mathbf{x}\|_2. \quad (35)$$

For each  $\tilde{\mathbf{V}}_{\mathbf{x}} \in \mathcal{A}$  we have, by definition, that  $\|\mathbf{x}\|_2 \leq 1$  which means that

$$d_F(\mathcal{A}) = 1, \quad \text{and} \quad d_{2 \rightarrow 2}(\mathcal{A}) \leq \sqrt{\frac{s}{N_t}}. \quad (36)$$

In order to estimate the entropy integral  $\mathcal{D}$  in (34), it is in particular necessary to provide good bounds for the appearing covering numbers  $\mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)$ . In view of (36) it is sufficient to consider  $0 < u \leq \sqrt{s/N_t}$ . The proof of Lemma 11 can be found in Appendix B.

**Lemma 11.** For  $0 < u \leq \sqrt{s/N_t}$ , it holds

$$\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \lesssim s \min \left\{ \log \left( \frac{N}{u^2 N_t} \right), \frac{1}{u^2 N_t} \log^2(N) \right\}. \quad (37)$$

Now we are able to estimate Dudley's integral as

$$\mathcal{D} \lesssim \sqrt{\frac{s}{N_t}} \log(eN) \log(s). \quad (38)$$

To this end, we split the integral into two pieces,

$$\begin{aligned} \mathcal{D} &= \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)} du \\ &\leq \int_0^{\frac{1}{\sqrt{N_t}}} \sqrt{\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)} du + \int_{\frac{1}{\sqrt{N_t}}}^{\frac{\sqrt{s}}{\sqrt{N_t}}} \sqrt{\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)} du =: \mathcal{I}_1 + \mathcal{I}_2. \end{aligned}$$

We estimate the first integral by using the first bound of Lemma 11 to obtain

$$\mathcal{I}_1 \lesssim \int_0^{\frac{1}{\sqrt{N_t}}} \sqrt{s \log \left( \frac{N}{u^2 N_t} \right)} du \leq \sqrt{s} \sqrt{\int_0^{\frac{1}{\sqrt{N_t}}} 1 du \int_0^{\frac{1}{\sqrt{N_t}}} \log \left( \frac{N}{u^2 N_t} \right) du},$$

where, in the second step, we used the Cauchy-Schwarz inequality. The latter integral can easily be calculated as

$$\begin{aligned} \int_0^{\frac{1}{\sqrt{N_t}}} \log \left( \frac{N}{u^2 N_t} \right) du &= \int_0^{\frac{1}{\sqrt{N_t}}} \log(N/N_t) - \log(u^2) du = \frac{1}{\sqrt{N_t}} \log(N/N_t) - \left[ u(\log(u^2) - 2) \right]_0^{1/\sqrt{N_t}} \\ &= \frac{1}{\sqrt{N_t}} (\log(N) + 2), \end{aligned}$$

which, together with the estimate from above, yields

$$\mathcal{I}_1 \lesssim \sqrt{\frac{s}{N_t}} \sqrt{\log(N) + 2}.$$

For the second integral  $\mathcal{I}_2$  we use the second bound provided by Lemma 11 and obtain

$$\mathcal{I}_2 \lesssim \int_{\frac{1}{\sqrt{N_t}}}^{\frac{\sqrt{s}}{\sqrt{N_t}}} \sqrt{\frac{s}{u^2 N_t} \log^2(N)} du = \sqrt{\frac{s}{N_t}} \log(N) \int_{\frac{1}{\sqrt{N_t}}}^{\frac{\sqrt{s}}{\sqrt{N_t}}} u^{-1} du = \sqrt{\frac{s}{N_t}} \log(N) \log(\sqrt{s}).$$

A combination of the estimates for  $\mathcal{I}_1$  and  $\mathcal{I}_2$  from above yields (38).

Now, having a bound for the entropy integral  $\mathcal{D}$ , we are well equipped to apply Theorem 10 and, thereby, prove Theorem 2. To this end let  $c > 0$  denote a constant to be chosen sufficiently large and

$$N_t \geq c\delta^{-2}s \max \{ \log^2(eN) \log^2(s), \log(1/\varepsilon) \}. \quad (39)$$

Due to (38) it follows that, for some constant  $C > 0$ ,

$$\mathcal{D} \leq C \sqrt{\frac{s}{N_t}} \log(eN) \log(s).$$

These two bounds imply that  $\mathcal{D} \leq C\delta/\sqrt{c}$  and, therefore, since  $d_F(\mathcal{A}) = 1$ , we obtain for the quantity  $E$  from Theorem 10 (if  $c$  is chosen sufficiently large),

$$E = \mathcal{D}(\mathcal{D} + 1) \leq (C\delta)^2/c + C\delta/\sqrt{c} < \frac{\delta}{2c_1},$$

where  $c_1$  is the constant from Theorem 10. A direct application of Theorem 10, again using  $d_F(\mathcal{A}) = 1$ , yields

$$\mathbb{P}(\delta_s \geq \delta) \leq \mathbb{P}(\delta_s \geq c_1 E + \delta/2) \leq 2 \exp \left( -\frac{c_2 \delta^2}{4d_{2 \rightarrow 2}^2(\mathcal{A})} \right) \leq 2 \exp \left( -\frac{c_2 \delta^2 N_t}{4s} \right),$$

where for the last inequality we used the estimate for  $d_{2 \rightarrow 2}(\mathcal{A})$  from (36). Due to (39), the latter term is bounded by  $\varepsilon$ , again provided that  $c$  is sufficiently large, which finalizes the proof of Theorem 2.  $\square$

### 3 Nonuniform recovery results

The proofs of our nonuniform recovery results (see Theorems 7 and 8) rely on the fact that — for sufficiently well balanced support sets  $S \subset \mathcal{G}$  — the singular values of the matrix  $\tilde{\mathbf{A}}_S$  are close to one with high probability. Along with further properties, this enables us to apply general recovery results for both the LASSO and basis pursuit denoising which we take from [6] and [14], respectively. In the following analysis we consider a scaled version of the considered measurement process, namely

$$\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{n}}, \quad \tilde{\mathbf{y}} = \frac{1}{\sqrt{N_T N_R N_t}} \mathbf{y}, \quad \tilde{\mathbf{A}} = \frac{1}{\sqrt{N_T N_R N_t}} \mathbf{A}, \quad \tilde{\mathbf{n}} = \frac{1}{\sqrt{N_T N_R N_t}} \mathbf{n}.$$

In the situation of Theorem 7,  $\tilde{\mathbf{n}}$  is now a mean-zero complex Gaussian random vector with variance  $\tilde{\sigma}^2 = \sigma^2 / N_T N_R N_t$ .

#### 3.1 Exact support recovery via the (debiased) LASSO

In order to prove Theorem 7 we use a general recovery result from [6], which provides the sufficient conditions  $(C_1)$ – $(C_5)$  below on fixed instances of the measurement matrix  $\tilde{\mathbf{A}}$ , the target scene vector  $\mathbf{x}$  and the noise vector  $\tilde{\mathbf{n}}$  that imply perfect reconstruction of the support set  $\text{supp}(\mathbf{x})$  from the measurements  $\tilde{\mathbf{y}}$  via the LASSO

$$\min_{\mathbf{x}} \frac{1}{2} \|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{y}}\|_2^2 + 2\mu \|\mathbf{x}\|_1, \quad \mu = \tilde{\sigma} \sqrt{2 \log(N)}. \quad (40)$$

Later (see Sections 3.1.2, 3.1.3 below) we will show that these conditions are fulfilled with high probability. In the following we write  $\Pi_S$  to denote the projection onto the linear space spanned by the columns of  $\tilde{\mathbf{A}}_S$ .

$(C_1)$  The matrix  $\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S$  is invertible and obeys  $\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1}\|_{2 \rightarrow 2} \leq 2$ ,

$(C_2)$   $\|\tilde{\mathbf{A}}_{S^c}^* \tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty < 1/4$ ,

$(C_3)$   $\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^* \tilde{\mathbf{n}}\|_\infty \leq 2\mu$ ,

$(C_4)$   $\|\tilde{\mathbf{A}}_{S^c}^* (\mathbf{Id} - \Pi_S) \tilde{\mathbf{n}}\|_\infty \leq \sqrt{2}\mu$ ,

$(C_5)$   $\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty \leq 3$ .

In [6] it is shown, using Conditions  $(C_1)$ – $(C_5)$ , that the difference vector  $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}$  between the solution  $\hat{\mathbf{x}}$  to the LASSO and the original vector  $\mathbf{x}$  fulfills  $\text{supp}(\mathbf{h}) \subset \text{supp}(\mathbf{x})$  so that  $\text{supp}(\hat{\mathbf{x}}) = \text{supp}(\mathbf{x})$  and, even more, the explicit formula (41) for  $\mathbf{h}$  is given.

**Lemma 12** (cf. [6, Lem. 3.4]). *Let  $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{n}}$  be given, and suppose that Conditions  $(C_1)$ – $(C_5)$  hold true, with  $S = \text{supp}(\mathbf{x})$ . If*

$$\min_{\theta \in S} |x_\theta| > 8\mu,$$

*then the solution to the LASSO (40) is given by  $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{h}$ , with*

$$\begin{aligned} \mathbf{h}_S &= (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{n}} - 2\mu \text{sgn}(\mathbf{x}_S)), \\ \mathbf{h}_{S^c} &= \mathbf{0}. \end{aligned} \quad (41)$$

##### 3.1.1 Estimates for the conditions $(C_1)$ – $(C_5)$

With Lemma 12 at hand we have a basic pattern for the proof Theorem 7. Indeed, one easily verifies that, assuming Conditions  $(C_1)$ – $(C_5)$  are fulfilled, Theorem 7 follows by rescaling the LASSO functional (19) with the factor  $N_T N_R N_t$  and recalling that  $\tilde{\sigma} = \sigma / \sqrt{N_T N_R N_t}$ . In order to show that Conditions  $(C_1)$ – $(C_5)$  hold true with high probability we conduct a probabilistic analysis of the extremal singular values of the matrix  $\tilde{\mathbf{A}}_S$ . Since columns of the matrix  $\tilde{\mathbf{A}}_S$  belonging to different angle classes (see Section

1.4 for the definition) are not correlated, it suffices to consider submatrices corresponding to the subsets  $S_{[\beta]} \subset S$  (see (24)) representing equivalent indices. In this sense, the following lemma is the main ingredient for the verification of conditions  $(C_1)$ – $(C_5)$ . The proof of Proposition 13 is rather long and, therefore, postponed until Section 3.3.

**Proposition 13.** *Let  $S \subset \mathcal{G}$  be a support set and let  $[\beta]$  be an angle class such that  $S_{[\beta]} \subset S$  is not empty. Then, for all  $\delta \in (0, 1)$ ,*

$$\mathbb{P}(\|\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}} - \mathbf{Id}\|_{2 \rightarrow 2} \geq \delta) \leq 7|S_{[\beta]}| \exp\left(-\frac{\delta\sqrt{N_t}}{8\sqrt{|S_{[\beta]}|}}\right).$$

Before we turn to the verification of Conditions  $(C_1)$ – $(C_5)$ , we derive two corollaries which follow directly from Proposition 13. We have already pointed out in (23) that, due to the fact that columns of  $\tilde{\mathbf{A}}$  belonging to different angle classes are not correlated, it holds

$$\mathbb{P}(\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \geq \delta) = \mathbb{P}\left(\max_{[\beta]} \|\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}} - \mathbf{Id}\|_{2 \rightarrow 2} \geq \delta\right).$$

Using the union bound and the tail bounds provided by Proposition 13 we arrive at

$$\mathbb{P}(\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \geq \delta) \leq 7 \sum_{[\beta]} |S_{[\beta]}| \exp\left(-\frac{\delta\sqrt{N_t}}{8\sqrt{|S_{[\beta]}|}}\right), \quad (42)$$

where the sum is taken over all angle classes  $[\beta]$  (with the property that  $S_{[\beta]}$  is not empty). Now it suffices to plug in the inequality  $|S_{[\beta]}| \leq \eta|S|/N_R$  which applies for  $\eta$ -balanced support sets to obtain the following corollary (recall, that there are exactly  $N_R$  angle classes  $[\beta]$ ).

**Corollary 14.** *For each  $\eta$ -balanced support set  $S \subset \mathcal{G}$  it holds*

$$\mathbb{P}(\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \geq \delta) \leq 7\eta|S| \exp\left(-\frac{\delta\sqrt{N_t N_R}}{8\sqrt{\eta|S|}}\right).$$

We use the estimate (42) to derive a bound on the coherence of  $\tilde{\mathbf{A}}$ . Note, that the coherence of this matrix fulfills

$$\max_{\Theta \neq \Theta'} |\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle| = \max_{[\beta]} \max_{\Theta, \Theta' \in S_{[\beta]}, \Theta \neq \Theta'} |\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle|.$$

Let  $\Theta, \Theta' \in S_{[\beta]}$  be given. For any matrix  $\mathbf{M}$  and each entry  $m_{i,j}$  it holds  $|m_{i,j}| \leq \|\mathbf{M}\|_{2 \rightarrow 2}$ . Now, since  $\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle$  is an entry of the matrix  $\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}$ , where we set  $S := \{\Theta, \Theta'\}$ , we obtain

$$\mathbb{P}(|\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle| \geq u) \leq 14 \exp\left(-\frac{u\sqrt{N_t}}{8\sqrt{2}}\right)$$

from (42). Since we can choose  $\binom{N/N_R}{2} \leq (N/N_R)^2/2$  pairs  $\Theta, \Theta'$  and there are  $N_R$  angle classes, an application of the union bound yields the following corollary.

**Corollary 15.** *The coherence of the matrix  $\tilde{\mathbf{A}}$  satisfies*

$$\mathbb{P}(\max_{\Theta \neq \Theta'} |\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle| \geq u) \leq \frac{7N^2}{N_R} \exp\left(-\frac{u\sqrt{N_t}}{8\sqrt{2}}\right).$$

Next we will derive upper bounds for the respective probabilities that one of the conditions  $(C_1)$ – $(C_5)$  fails to hold.

### Condition $(C_1)$

We define events  $\mathcal{C}_1^\delta$  as

$$\mathcal{C}_1^\delta = \{\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq \delta \leq 1/2\}.$$

Clearly, if  $\mathcal{C}_1^\delta$  occurs (i.e.,  $\delta \in (0, 1/2]$ ), then it holds  $\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1}\|_{2 \rightarrow 2} \leq 1/(1 - \delta) \leq 2$ . Hence, Corollary 14 can be used to show that condition  $(C_1)$  is fulfilled with high probability.



**Condition (C<sub>2</sub>)**

**Lemma 16.** *Let  $S \subset \mathcal{G}$  be a  $\eta$ -balanced support set. If there exist  $\delta \in (0, 1)$  and  $u > 0$  such that*

$$\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq \delta, \quad \max_{\Theta \neq \Theta'} |\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle| \leq u, \quad (43)$$

*then it holds*

$$\mathbb{P}(\|\tilde{\mathbf{A}}_{S^c}^* \tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty \geq 1/4) \leq 2N \exp\left(-\frac{(1-\delta)^2 N_R}{32u^2 \eta |S|}\right).$$

*Proof.* For  $\Theta \in S^c$  set  $\mathbf{v}_\Theta := (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_\Theta$ . Since  $\text{sgn}(\mathbf{x}_S)$  forms a Steinhaus sequence, the Hoeffding-type inequality for Steinhaus sequences of Lemma 31 yields

$$\mathbb{P}(|\langle \mathbf{v}_\Theta, \text{sgn}(\mathbf{x}_S) \rangle| \geq 1/4) \leq 2e^{-1/(32\|\mathbf{v}_\Theta\|_2^2)}. \quad (44)$$

In order to derive a bound for the norms  $\|\mathbf{v}_\Theta\|_2$  we calculate,

$$\|\mathbf{v}_\Theta\|_2 = \|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_\Theta\|_2 \leq \|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1}\|_{2 \rightarrow 2} \sqrt{\sum_{\tilde{\Theta} \in S} |\langle \tilde{\mathbf{A}}_{\tilde{\Theta}}, \tilde{\mathbf{A}}_\Theta \rangle|^2}.$$

By assumption,  $\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1}\|_{2 \rightarrow 2}$  is bounded by  $(1-\delta)^{-1}$ . Since  $S$  is  $\eta$ -balanced, there are at most  $\eta|S|/N_R$  indices  $\tilde{\Theta}$  such that the summands are not equal to zero (recall that columns corresponding to different angle classes are not correlated). Together with the fact that each such inner product  $|\langle \tilde{\mathbf{A}}_{\tilde{\Theta}}, \tilde{\mathbf{A}}_\Theta \rangle|$  is bounded by  $u$ , we have  $\|\mathbf{v}_\Theta\|_2 \leq u\sqrt{\eta|S|}/(1-\delta)\sqrt{N_R}$ . Inserting this estimate into (44) and furthermore applying the union bound yields the assertion.  $\square$

**Condition (C<sub>5</sub>)**

**Lemma 17.** *If the matrix  $\tilde{\mathbf{A}}_S$  obeys  $\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq \delta < 1$ , and  $\text{sgn}(\mathbf{x}_S)$  is a (random) Steinhaus vector, then it holds*

$$\mathbb{P}(\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty > 3) \leq 2|S|e^{-2(1-\delta)^2/\delta^2}.$$

*Proof.* An application of the triangle inequality yields

$$\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty \leq \|\text{sgn}(\mathbf{x}_S)\|_\infty + \|((\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} - \mathbf{Id}) \text{sgn}(\mathbf{x}_S)\|_\infty.$$

Since the first summand on the right-hand side is equal to 1, it suffices to show that the second term is bounded by 2 with high probability. Let  $\mathbf{v}_\Theta$ ,  $\Theta \in S$ , denote the rows of the matrix  $(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} - \mathbf{Id}$ . Due to the assumption it holds  $\|\mathbf{Id} - \tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S\|_{2 \rightarrow 2} \leq \delta$ . Thus,

$$\|\mathbf{v}_\Theta\|_2 = \|((\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} - \mathbf{Id})\mathbf{e}_\Theta\|_2 = \left\| \sum_{k=1}^{\infty} (\mathbf{Id} - \tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^k \mathbf{e}_\Theta \right\|_2 \leq \frac{1}{1-\delta} \|\mathbf{Id} - \tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S\|_{2 \rightarrow 2} \leq \frac{\delta}{1-\delta}.$$

Using Hoeffding's inequality for Steinhaus sums (Lemma 31) we obtain

$$\mathbb{P}(|\langle \mathbf{v}_\Theta, \text{sgn}(\mathbf{x}_S) \rangle| \geq 2) \leq 2e^{-2/\|\mathbf{v}_\Theta\|_2^2} \leq 2e^{-2(1-\delta)^2/\delta^2}.$$

Finally, the assertion follows by applying the union bound.  $\square$

**Condition (C<sub>3</sub>)**

**Lemma 18.** *If the matrix  $\tilde{\mathbf{A}}_S$  obeys  $\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq 1/2$ , then, for a mean-zero Gaussian vector  $\tilde{\mathbf{n}}$  with independent entries of variance  $\tilde{\sigma}^2$  and  $\mu = \tilde{\sigma}\sqrt{2\log(N)}$ , it holds,*

$$\mathbb{P}_{\mathbf{n}}(\|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^* \tilde{\mathbf{n}}\|_\infty \geq 2\mu) \leq N^{-3}.$$

*Proof.* For  $\Theta \in S$  we write  $\mathbf{v}_\Theta$  to denote the row of the matrix  $(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^*$  corresponding to the index  $\Theta$ . An application of Lemma 29 gives, since  $\tilde{\sigma}^{-1} \tilde{\mathbf{n}}$  is a standard complex Gaussian random vector,

$$\mathbb{P}_{\mathbf{n}}(|\langle \mathbf{v}_\Theta, \tilde{\mathbf{n}} \rangle| \geq 2\mu) = \mathbb{P}_{\mathbf{n}}(|\langle \mathbf{v}_\Theta, \tilde{\sigma}^{-1} \tilde{\mathbf{n}} \rangle| \geq 2\sqrt{2\log(N)}) \leq e^{-8\log(N)/\|\mathbf{v}_\Theta\|_2^2}.$$

Since  $\mathbf{v}_\Theta$ ,  $\Theta \in S$ , is a column of the matrix  $\mathbf{M} := \tilde{\mathbf{A}}_S(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1}$ , the operator norm of  $\mathbf{M}$  provides a uniform bound for the  $\ell_2$ -norms of the rows  $\mathbf{v}_\Theta$ . By assumption  $\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq 1/2$ , and, hence,

$$\|\mathbf{v}_\Theta\|_2^2 \leq \|\mathbf{M}\|_{2 \rightarrow 2}^2 = \|\mathbf{M}^* \mathbf{M}\|_{2 \rightarrow 2} = \|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1}\|_{2 \rightarrow 2} \leq (1 - 1/2)^{-1} = 2.$$

Combining this with the tail bound above yields

$$\mathbb{P}_{\mathbf{n}}(|\langle \mathbf{v}_\Theta, \tilde{\mathbf{n}} \rangle| \geq 2\mu) \leq e^{-8\log(N)/2} = N^{-4}.$$

An application of the union bound, using  $|S| \leq N$ , implies the assertion.  $\square$

#### Condition (C<sub>4</sub>)

In order to deal with Condition (C<sub>4</sub>), we need an estimate in the  $\ell_2$ -norms of the columns of the matrix  $\tilde{\mathbf{A}}$ .

**Lemma 19.** *The  $\ell_2$ -norms of the columns of the matrix  $\tilde{\mathbf{A}}$  satisfy*

$$\mathbb{P}(\max_{\Theta \in \mathcal{G}} \|\tilde{\mathbf{A}}_\Theta\|_2 \geq 2/\sqrt{3}) \leq N e^{-(2/\sqrt{3}-1)^2 N_t}.$$

*Proof.* It follows from the definition of the columns  $\tilde{\mathbf{A}}$  in (8) that  $\|\tilde{\mathbf{A}}_\Theta\|_2$ ,  $\Theta \in \mathcal{G}$ , is given by  $\frac{1}{\sqrt{N_t}} \|\mathbf{z}\|_2$  with  $\mathbf{z} = \frac{1}{\sqrt{N_t}} \sum_{i=1}^{N_t} e^{i2\pi \cdot d_T \beta \Delta_\beta(i-1)} \mathbf{s}_i$ , where the  $\mathbf{s}_i$  are independent standard complex Gaussian vectors. Thus, also  $\mathbf{z}$  is an  $N_t$ -dimensional standard complex Gaussian random vector so that Lemma 30 yields

$$\mathbb{P}(\|\tilde{\mathbf{A}}_\Theta\|_2 \geq 2/\sqrt{3}) = \mathbb{P}(\|\mathbf{z}\|_2 \geq \sqrt{N_t} + (2/\sqrt{3} - 1)\sqrt{N_t}) \leq e^{-(2/\sqrt{3}-1)^2 N_t}.$$

The assertion follows by applying the union bound.  $\square$

Under the assumption that the columns are bounded in the  $\ell_2$ -norm, we are able to provide an estimate for the probability that condition (C<sub>4</sub>) fails to hold.

**Lemma 20.** *Let  $\tilde{\mathbf{n}}$  be a mean-zero Gaussian vector with covariance matrix  $\tilde{\sigma}^2 \mathbf{Id}$  and  $\mu = \tilde{\sigma} \sqrt{2\log(N)}$ . If  $\|\tilde{\mathbf{A}}_\Theta\|_2 \leq 2/\sqrt{3}$  for all  $\Theta \in S^c$ , then*

$$\mathbb{P}_{\mathbf{n}}(\|\tilde{\mathbf{A}}_{S^c}^* (\mathbf{Id} - \Pi_S) \tilde{\mathbf{n}}\|_\infty \geq \sqrt{2}\mu) \leq N^{-2}.$$

*Proof.* Note that the  $\ell_2$ -norms of the rows of the matrix  $\tilde{\mathbf{A}}_{S^c}^* (\mathbf{Id} - \Pi_S)$  are uniformly bounded by

$$\max_{\Theta \in S^c} \|(\mathbf{Id} - \Pi_S) \tilde{\mathbf{A}}_\Theta\|_2 \leq \underbrace{\|\mathbf{Id} - \Pi_S\|_{2 \rightarrow 2}}_{\leq 1} \max_{\Theta \in S^c} \|\tilde{\mathbf{A}}_\Theta\|_2 \leq 2/\sqrt{3}.$$

Using the same arguments as in the proof of Lemma 18 (now with  $\mathbf{v}_\Theta$ ,  $\Theta \in S^c$ , denoting the columns of the matrix  $(\mathbf{Id} - \Pi_S) \tilde{\mathbf{A}}_\Theta$ ), we obtain

$$\mathbb{P}_{\mathbf{n}}(|\langle \mathbf{v}_\Theta, \tilde{\mathbf{n}} \rangle| \geq \sqrt{2}\mu) \leq e^{-4\log(N)/\|\mathbf{v}_\Theta\|_2^2} \leq e^{-3\log(N)} = N^{-3},$$

where, in the last step, we used the bound on the norm  $\|\mathbf{v}_\Theta\|_2$  from above. An application of the union bound, using that  $|S^c| \leq N$ , implies the assertion.  $\square$

### 3.1.2 Proof of Theorem 7(a)

Now we are prepared to finalize the proof of Theorem 7. In view of Lemma 12, we only have to verify that Conditions  $(C_1)$ – $(C_5)$  are fulfilled with probability exceeding  $1 - 7 \max\{\varepsilon, N^{-2}\}$  in order to show part (a) of Theorem 7. As we have already noted, Condition  $(C_1)$  holds, if  $\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq \delta$ , for some  $\delta \in (0, 1/2]$ . Thus, Corollary 14 can be used to show that  $(C_1)$  is fulfilled with high probability. Moreover, a careful inspection of Corollaries 14, 15 and Lemmas 16–20 reveals that the probability of the assertion of Theorem 7 *failing* to hold true is bounded by the probability  $\mathbb{P}(\neg \mathcal{A}_{\delta,u}) + \sum_{k=2}^5 \mathbb{P}(\neg \mathcal{C}_k | \mathcal{A}_{\delta,u})$ , where each  $\mathcal{C}_k$  stands for the event when Condition  $(C_k)$  is fulfilled, and the event  $\mathcal{A}_{\delta,u}$  is defined as

$$\mathcal{A}_{\delta,u} = \underbrace{\{\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} < \delta\}}_{=: \mathcal{C}_1^\delta} \cap \underbrace{\{\max_{\Theta \neq \Theta'} |\langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle| \leq u\}}_{=: \mathcal{E}_1^u} \cap \underbrace{\{\max_{\Theta \in \mathcal{G}} \|\tilde{\mathbf{A}}_\Theta\|_2 \leq 2/\sqrt{3}\}}_{=: \mathcal{E}_2}. \quad (45)$$

Now, Corollaries 14, 15 and Lemma 19 imply

$$\begin{aligned} \mathbb{P}(\neg \mathcal{A}_{\delta,u}) &\leq \mathbb{P}(\neg \mathcal{C}_1^\delta) + \mathbb{P}(\neg \mathcal{E}_1^u) + \mathbb{P}(\neg \mathcal{E}_2) \\ &\leq 7\eta|S| \exp\left(-\frac{\delta\sqrt{N_R N_t}}{8\sqrt{\eta}|S|}\right) + \frac{7N^2}{N_R} \exp\left(-\frac{u\sqrt{N_t}}{8\sqrt{2}}\right) + N e^{(2/\sqrt{3}-1)^2 N_t}. \end{aligned} \quad (46)$$

Due Lemmas 16, 17, 18, and 20 we have the following bounds on the failure probability of Conditions  $(C_2)$ ,  $(C_5)$ ,  $(C_3)$ , and  $(C_4)$  — conditional on the event  $\mathcal{A}_{\delta,u}$ :

$$\begin{aligned} \mathbb{P}(\neg \mathcal{C}_2 | \mathcal{A}_{\delta,u}) &\leq 2N e^{-(1-\delta)^2 N_R / 32 u^2 \eta |S|}, & \mathbb{P}(\neg \mathcal{C}_5 | \mathcal{A}_{\delta,u}) &\leq 2|S| e^{-2(1-\delta)^2 / \delta^2}, \\ \mathbb{P}(\neg \mathcal{C}_3 | \mathcal{A}_{\delta,u}) &\leq N^{-3}, & \mathbb{P}(\neg \mathcal{C}_4 | \mathcal{A}_{\delta,u}) &\leq N^{-2}. \end{aligned} \quad (47)$$

Since we claim that the assertion of Theorem 7 holds true with probability exceeding  $1 - 7 \max\{\varepsilon, N^{-2}\}$ , it is enough to verify that each of the seven terms in (46), (47), (48) is either bounded by  $\varepsilon$  or by  $N^{-2}$ . Since this holds trivially for the terms in (48), we are left with the terms in (46) and (47) which, as we will see, can all be bounded by  $\varepsilon$ . Due to the definition of the parameter  $\eta$  (cf. Definition 6) it holds (note that each equivalence class contains  $N/N_R$  elements)

$$\eta = \max_{[\beta]} \frac{|S_{[\beta]}|}{|S|/N_R} \leq \frac{N/N_R}{|S|/N_R} = \frac{N}{|S|}.$$

Thus,  $\eta|S| \leq N$ , which yields that in order to verify that the first term in (46) is smaller or equal to  $\varepsilon$  it is sufficient to show that

$$\frac{64\eta|S|}{\delta^2} \log^2\left(\frac{7N}{\varepsilon}\right) \leq N_R N_t, \quad (49)$$

whereas for the second term in (46) to be smaller or equal to  $\varepsilon$  it suffices that

$$\frac{128N_R}{u^2} \log^2\left(\frac{7N^2}{\varepsilon N_R}\right) \leq N_R N_t. \quad (50)$$

The term  $\log(7N^2/\varepsilon N_R)$  is dominated by  $2 \log(7N/\varepsilon)$  and, therefore, we can set

$$u := \delta \sqrt{8N_R} / \sqrt{\eta|S|} \quad (51)$$

in order to ensure that (50) is already implied by (49). In order to show (49), we set

$$\delta = \frac{1}{\sqrt{K \log(eN/\varepsilon)}},$$

where  $K > 0$  will be chosen below. With this the left-hand side of (49) can be estimated as

$$\frac{64\eta|S|}{\delta^2} \log^2\left(\frac{7N}{\varepsilon}\right) = 64K\eta|S| \log\left(\frac{eN}{\varepsilon}\right) \log^2\left(\frac{7N}{\varepsilon}\right) \leq 64K \log^3(7)\eta|S| \log^3\left(\frac{N}{\varepsilon}\right).$$

Due to the assumption (25) we have

$$N_R N_t \geq C \eta |S| \log^3 \left( \frac{N}{\varepsilon} \right),$$

where  $C > 0$  is a suitable constant. By setting  $C = 64K \log^3(7)$ , we establish (49) (and, hence, also (50)) which means that both the first and the second term in (46) are bounded by  $\varepsilon$ . In order to ensure that the last term  $N e^{-(2/\sqrt{3}-1)^2 N_t}$  in (46) is bounded by the second one (and, hence, is also smaller or equal to  $\varepsilon$ ), it suffices to have  $(2/\sqrt{3}-1)^2 N_t \geq u \sqrt{N_t}/8\sqrt{2}$  which, recalling the definition of  $u$ , is equivalent to

$$\delta \leq 4(2/\sqrt{3}-1)^2 \sqrt{\frac{\eta |S| N_t}{N_R}}.$$

Since by definition of the parameter  $\eta$  we have  $\eta |S|/N_R \geq |S_{[\beta]}| \geq 1$  and, moreover,  $N_t \geq 1$ , anyway, this latter condition is fulfilled as long as  $\delta \leq 4(2/\sqrt{3}-1)^2$  which is fulfilled whenever  $K$  is sufficiently large. Since  $\log(eN/\varepsilon) \geq 1$ , and, hence,  $\delta \leq 1/\sqrt{K}$ , we have for the first term in (47), recalling that  $u^2 = 8\delta^2 N_R/\eta |S|$ ,

$$\begin{aligned} 2N \exp \left( -\frac{[1-\delta]^2 N_R}{32u^2 \eta |S|} \right) &= 2N \exp \left( -\frac{[1-\delta]^2}{256\delta^2} \right) \\ &= 2N \exp \left( -\frac{1}{256} \left[ 1 - \frac{1}{\sqrt{K \log(eN/\varepsilon)}} \right]^2 K \log \left( \frac{eN}{\varepsilon} \right) \right) \\ &\leq 2N \exp \left( -\frac{1}{256} \left[ 1 - \frac{1}{\sqrt{K}} \right]^2 K \log \left( \frac{2N}{\varepsilon} \right) \right) = 2N \exp \left( -\frac{1}{256} [\sqrt{K} - 1]^2 \log \left( \frac{2N}{\varepsilon} \right) \right). \end{aligned}$$

Therefore, in order to show that the first term in (47) is bounded by  $\varepsilon$ , it is sufficient to ensure that  $(\sqrt{K}-1)^2 \geq 256 = 16^2$  which holds true if  $\sqrt{K} \geq 17$ . The second term in (47) is bounded by the first one. Finally, the condition  $\delta \leq 1/2$  is satisfied since  $\delta \leq 1/\sqrt{K} \leq 1/17$ . This finishes the proof Theorem 7(a).  $\square$

### 3.1.3 Proof of Theorem 7(b)

For part (b) it remains to derive the stated approximation property of the solution to the debiased LASSO with respect to  $\mathbf{x}_S$ . From the proof of part (a) it follows that, with high probability (as stated in the theorem), the columns of the matrix  $\mathbf{A}_S$  are linearly independent. Hence, the solution  $\mathbf{z}^\#$  to the least squares problem (20) is given by  $\mathbf{z}^\# = (\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^* \mathbf{y}$ . Indeed, this formula can easily be verified by calculating the corresponding optimality conditions. Since  $\mathbf{x}$  is exactly sparse and supported on the set  $S$  it holds  $\mathbf{y} = \mathbf{A}_S \mathbf{x}_S + \mathbf{n}$ , i.e.,

$$\mathbf{z}^\# = \mathbf{x}_S + (\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^* \mathbf{n}.$$

Now, by a standard inequality between the  $\ell_2$  and the  $\ell_\infty$ -norm,

$$\|\mathbf{z}^\# - \mathbf{x}_S\|_2 = \|(\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^* \mathbf{n}\|_2 \leq \sqrt{s} \|(\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^* \mathbf{n}\|_\infty. \quad (52)$$

Since we may assume that we are in the event where condition (C<sub>3</sub>) holds true, we have the estimate

$$\|(\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^* \mathbf{n}\|_\infty = \|(\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^* \tilde{\mathbf{n}}\|_\infty \leq 2\mu = 2\tilde{\sigma} \sqrt{2 \log(N)} = \frac{2\sigma}{\sqrt{N_T N_R N_t}} \sqrt{2 \log(N)},$$

where, in the second but last step, we plugged in the definition of  $\mu$  from (40). A combination of this latter inequality with (52) implies the desired approximation property.  $\square$

### 3.2 Recovery via basis pursuit denoising under sparsity defect

We replace the original constraint  $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \varrho$  of the basis pursuit denoising program (11) by the equivalent constraint  $\|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{y}}\|_2 \leq \tilde{\varrho}$ , where  $\tilde{\varrho} = \varrho/\sqrt{N_T N_R N_t}$ . For the proof of Theorem 8 we use a general recovery result for basis pursuit denoising stated as Theorem 21 below. The original result [14, Thm. 4.33] uses an inexact dual certificate which in our case can actually be taken as the canonical exact dual certificate given by

$$\tilde{\mathbf{h}} = \tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S).$$

Hence, the following Theorem 21, as it is formulated here, is a special case of [14, Thm. 4.33], see also [19]. The constants  $C_1$  and  $C_2$  appearing in (54) below can be calculated explicitly using the corresponding formulas in [14, Thm. 4.33]. However, we refrain from doing so at this point.

**Theorem 21** (cf. [14, Thm. 4.33]). *Let  $\mathbf{x} \in \mathbb{C}^N$  be given with  $s$  largest absolute entries supported on  $S \subset [N]$ . Assume that measurements  $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{n}} \in \mathbb{C}^m$  are given with  $m < N$  and  $\|\tilde{\mathbf{n}}\|_2 \leq \tilde{\varrho}$ . If*

$$\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq \delta < 1, \quad \|\tilde{\mathbf{A}}_{S^c}^* \tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty < 1,$$

and, moreover, for  $\alpha_1, \alpha_2 \geq 0$ ,

$$\max_{\ell \in S^c} \|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_\ell\|_2 \leq \alpha_1, \quad \|\tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_2 \leq \alpha_2 \sqrt{s}, \quad (53)$$

then each solution  $\mathbf{z}^\#$  to the basis pursuit denoising program

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|\tilde{\mathbf{A}}\mathbf{z} - \tilde{\mathbf{y}}\|_2 \leq \tilde{\varrho}$$

satisfies

$$\|\mathbf{z}^\# - \mathbf{x}\|_2 \leq C_1 \inf_{s\text{-sparse } \mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|_1 + C_2 \sqrt{s} \tilde{\varrho}, \quad (54)$$

for some constants  $C_1$  and  $C_2$  depending only on  $\delta, \alpha_1, \alpha_2$ .

**Remark 22.** The error estimate (54) is slightly weaker as the one that holds under the RIP, see (12). In fact, (54) features the  $\ell_2$ -norm on the right hand side, but (12) states an error estimate in the  $\ell_1$ -norm. This seems the price one has to pay when not working with strong recovery conditions such as the RIP.

#### 3.2.1 Proof of Theorem 8

In order to prove Theorem 8 we show that the assumptions of Theorem 21 are fulfilled with high probability. This can be achieved by simply repeating some of the arguments we already used during the proof of Theorem 7. Indeed, recalling the event  $\mathcal{A}_{\delta,u}$  (see (45)) from the proof of Theorem 7, where we now choose  $\delta = 1/2$  and  $u$  as in (51), then it holds (for details see the proof of Theorem 8)

$$\mathcal{A}_{\delta,u} \subset \{\|\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S - \mathbf{Id}\|_{2 \rightarrow 2} \leq 1/2\} \cap \{\max_{\Theta \in \mathcal{G}} \|\tilde{\mathbf{A}}_\Theta\|_2 \leq 2/\sqrt{3}\}.$$

This means that, in case of the event  $\mathcal{A}_{\delta,u}$  it holds

$$\max_{\Theta \in S^c} \|\mathbf{A}_S^* \mathbf{A}_\Theta\|_2 \leq \|\mathbf{A}_S^*\|_{2 \rightarrow 2} \max_{\Theta \in S^c} \|\mathbf{A}_\Theta\|_2 \leq 2\sqrt{1+1/2}/\sqrt{3} = \sqrt{2},$$

and, moreover,

$$\|\mathbf{A}_S (\mathbf{A}_S^* \mathbf{A}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_2 \leq \|\mathbf{A}_S\|_{2 \rightarrow 2} \|(\mathbf{A}_S^* \mathbf{A}_S)^{-1}\|_{2 \rightarrow 2} \|\text{sgn}(\mathbf{x}_S)\|_2 \leq \frac{\sqrt{1+\delta}}{1-\delta} \sqrt{s} = \sqrt{6} \sqrt{s}.$$

As a consequence, by setting  $\alpha_1 = \sqrt{2}$  and  $\alpha_2 = \sqrt{6}$  the assumptions in (53) are automatically fulfilled, provided that the event  $\mathcal{A}_{\delta,u}$  occurs. Finally, as the proof of Theorem 7 shows, it holds  $\mathbb{P}(\neg \mathcal{A}_{\delta,u}) \leq 3\varepsilon$  and (by Lemma 16)

$$\mathbb{P}(\|\tilde{\mathbf{A}}_{S^c}^* \tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty \geq 1/4 \mid \mathcal{A}_{\delta,u}) \leq \varepsilon.$$

This is also a consequence of the special choice of  $u$ , depending on  $\delta$ , and the condition (28) on the number of targets  $s$  from the assumptions of Theorem 8. Thus, using the union bound, the probability that the assumptions of Theorem 21 fail to hold are bounded by

$$\mathbb{P}(\neg \mathcal{A}_{\delta,u}) + \mathbb{P}(\|\tilde{\mathbf{A}}_{S^c}^* \tilde{\mathbf{A}}_S (\tilde{\mathbf{A}}_S^* \tilde{\mathbf{A}}_S)^{-1} \text{sgn}(\mathbf{x}_S)\|_\infty \geq 1/4 \mid \mathcal{A}_{\delta,u}) \leq 3\varepsilon + \varepsilon = 4\varepsilon. \quad \square$$

### 3.3 Proof of Proposition 13

We now show Proposition 13 concerning the well-conditionedness of the submatrix  $\tilde{\mathbf{A}}_{S_{[\beta]}}$ . In what follows we write  $s_{(i,a)}$  for the  $a$ -th entry of the signal vector  $\mathbf{s}_i$ . In order to analyze the quantity  $\|\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}} - \mathbf{Id}\|_{2 \rightarrow 2}$  we write the matrix product  $\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}}$  as (see Appendix C, (76) for a proof)

$$\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}} = \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \overline{s_{(i,a)}} s_{(j,b)} \mathbf{Y}^{(i,a),(j,b)}, \quad (55)$$

where  $\mathbf{Y}^{(i,a),(j,b)}$  is an  $|S_{[\beta]}| \times |S_{[\beta]}|$  matrix which, for indices  $\Theta, \Theta' \in S_{[\beta]}$ ,  $\Theta = (\beta, \tau, f)$ ,  $\Theta' = (\beta', \tau', f')$ , is given entrywise by

$$[\mathbf{Y}^{(i,a),(j,b)}]_{\Theta, \Theta'} = \delta_{a-b, \tau' - \tau}^{\sim N_t} (N_T N_t)^{-1} e^{i2\pi \cdot \frac{f' - f}{N_t} (\tau + a - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta' (j-1) - \beta (i-1)]}, \quad (56)$$

and where  $\delta_{\cdot, \cdot}^{\sim N_t}$  is defined in (29). By subdividing the sum in (55) into an off-diagonal part containing the summands with  $(i,a) \neq (j,b)$  and a remaining diagonal part, and then applying the triangle inequality, it follows that

$$\begin{aligned} & \|\tilde{\mathbf{A}}_{S_{[\beta]}}^* \tilde{\mathbf{A}}_{S_{[\beta]}} - \mathbf{Id}\|_{2 \rightarrow 2} \\ & \leq \left\| \left[ \sum_{i=1}^{N_T} \sum_{a=1}^{N_t} |s_{(i,a)}|^2 \mathbf{Y}^{(i,a),(i,a)} \right] - \mathbf{Id} \right\|_{2 \rightarrow 2} + \left\| \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \overline{s_{(i,a)}} s_{(j,b)} \mathbf{Y}^{(i,a),(j,b)} \right\|_{2 \rightarrow 2} \\ & \quad \text{where } (i,a) \neq (j,b) \\ & =: \|\mathbf{Y}_= - \mathbf{Id}\|_{2 \rightarrow 2} + \|\mathbf{Y}_\neq\|_{2 \rightarrow 2}. \end{aligned} \quad (57)$$

**A tail bound for  $\|\mathbf{Y}_= - \mathbf{Id}\|_{2 \rightarrow 2}$**

By plugging in the definition of the entries  $[\mathbf{Y}^{(i,a),(i,a)}]_{\Theta, \Theta'}$  from (56) one finds

$$\sum_{i=1}^{N_T} \sum_{a=1}^{N_t} [\mathbf{Y}^{(i,a),(i,a)}]_{\Theta, \Theta'} = \delta_{\tau', \tau} (N_T N_t)^{-1} e^{i2\pi \cdot \frac{f' - f}{N_t} \tau} \sum_{a=1}^{N_t} e^{i2\pi \cdot \frac{f' - f}{N_t} (a-1)} \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T \Delta_\beta (\beta' - \beta)(i-1)} = \delta_{\Theta', \Theta},$$

i.e., the sum over all matrices  $\mathbf{Y}^{(i,a),(i,a)}$  is the identity matrix. Therefore, we can write the matrix  $\mathbf{Y}_= - \mathbf{Id}$  in (57) as a sum over random matrices, namely

$$\mathbf{Y}_= - \mathbf{Id} = \sum_{i=1}^{N_T} \sum_{a=1}^{N_t} \hat{\mathbf{Y}}^{(i,a)}, \quad \hat{\mathbf{Y}}^{(i,a)} = \chi_{(i,a)} \mathbf{Y}^{(i,a),(i,a)}, \quad \chi_{(i,a)} = |s_{(i,a)}|^2 - 1. \quad (58)$$

The random matrices  $\hat{\mathbf{Y}}^{(i,a)}$  have block diagonal structure. To see this, we partition the index set  $S_{[\beta]}$  as

$$S_{[\beta]} = \bigcup_{\tau \in [N_t]} S_{[\beta]}^\tau, \quad S_{[\beta]}^\tau = \{(\beta', \tau', f') \in S_{[\beta]} : \tau' = \tau\}.$$

Recall that the matrix  $\mathbf{Y}^{(i,a),(i,a)}$  is given entrywise by

$$[\mathbf{Y}^{(i,a),(i,a)}]_{\Theta, \Theta'} = \delta_{\tau, \tau'} (N_T N_t)^{-1} e^{i2\pi \cdot \frac{f' - f}{N_t} (\tau + a - 1)} e^{i2\pi \cdot d_T \Delta_\beta (\beta' - \beta)(i-1)}.$$

The appearance of the factor  $\delta_{\tau,\tau'}$  implies that the matrices  $\mathbf{Y}^{(i,a),(i,a)}$  (and, hence, also  $\widehat{\mathbf{Y}}^{(i,a)}$ ) have a block diagonal structure. In case of the matrices  $\widehat{\mathbf{Y}}^{(i,a)}$  one obtains — enumerating the indices  $(\beta, \tau, f) \in S_{[\beta]}$  adequately —

$$\widehat{\mathbf{Y}}^{(i,a)} = \chi_{(i,a)} \begin{bmatrix} \mathbf{\Lambda}_1^{(i,a)} & & & \\ & \mathbf{\Lambda}_2^{(i,a)} & & \\ & & \ddots & \\ & & & \mathbf{\Lambda}_{N_t}^{(i,a)} \end{bmatrix}, \quad \mathbf{\Lambda}_\tau^{(i,a)} = \mathbf{v}_\tau^{(i,a)} [\mathbf{v}_\tau^{(i,a)}]^*, \quad (59)$$

where the vectors  $\mathbf{v}_\tau^{(i,a)} \in \mathbb{C}^{|S_{[\beta]}^\tau|}$  are given entrywise as

$$[\mathbf{v}_\tau^{(i,a)}]_{(\beta,\tau,f)} = (N_T N_t)^{-1/2} e^{i2\pi \cdot \frac{f}{N_t}(\tau+a-1)} e^{i2\pi \cdot d_T \Delta_\beta \beta(i-1)}, \quad (\beta, \tau, f) \in S_{[\beta]}^\tau.$$

Note, since for particular choices of  $\tau \in [N_t]$  the index sets  $S_{[\beta]}^\tau$  might be empty, it is possible that some blocks  $\mathbf{\Lambda}_\tau^{(i,a)}$  appearing in (59) might be zero-dimensional. In order to analyze the random variables in (58), we use the following result, taken from [30]. Below we write  $\lambda_{\max}(\cdot)$  to denote the maximal eigenvalue of a given (self-adjoint) matrix.

**Theorem 23** (cf. [30, Thm. 6.2]). *Consider a finite sequence  $\{\mathbf{X}_k\}$  of independent, random, self-adjoint  $d \times d$  matrices. Assume that*

$$\mathbb{E}\mathbf{X}_k = 0 \quad \text{and} \quad \mathbb{E}\mathbf{X}_k^p \preceq \frac{p!}{2} R^{p-2} \mathbf{A}_k^2, \quad \text{for } p = 2, 3, 4, \dots \quad (60)$$

Compute the variance parameter

$$\sigma^2 = \left\| \sum_k \mathbf{A}_k^2 \right\|_{2 \rightarrow 2}. \quad (61)$$

Then the following chain of inequalities holds for all  $t \geq 0$ :

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_k \mathbf{X}_k\right) \geq t\right) \leq d \exp\left(-\frac{t^2/2}{\sigma^2 + Rt}\right) \leq \begin{cases} d \exp(-t^2/4\sigma^2), & \text{for } t \leq \sigma^2/R, \\ d \exp(-t/4R), & \text{for } t \geq \sigma^2/R. \end{cases} \quad (62)$$

In our case the matrices  $\mathbf{X}_k$  are given by the matrices  $\widehat{\mathbf{Y}}^{(i,a)}$  being defined in (59). Clearly these matrices are self-adjoint and, moreover  $\mathbb{E}\widehat{\mathbf{Y}}^{(i,a)} = 0$ . In order to provide for all further inequalities in (60) we first calculate powers of the block matrices  $\mathbf{\Lambda}_\tau^{(i,a)}$  in (59):

$$[\mathbf{\Lambda}_\tau^{(i,a)}]^p = \langle \mathbf{v}_\tau^{(i,a)}, \mathbf{v}_\tau^{(i,a)} \rangle^{p-2} [\mathbf{\Lambda}_\tau^{(i,a)}]^2 = \left(\frac{|S_{[\beta]}^\tau|}{N_T N_t}\right)^{p-2} [\mathbf{\Lambda}_\tau^{(i,a)}]^2. \quad (63)$$

The random variables  $\chi_{(i,a)}$  in (59) are subexponential with

$$\mathbb{P}(|\chi_{(i,a)}| \geq t) = \mathbb{P}(|s_{(i,a)}|^2 - 1| \geq t) \leq \mathbb{P}(|s_{(i,a)}|^2 \geq t-1) \leq e^{-(t-1)} = ee^{-t},$$

where the second inequality follows from Lemma 28. Using basically Fubini's Theorem and a change of variables (see, e.g., [14, Prop. 7.1]) and, furthermore, plugging in the tail estimate from above, we obtain for  $p \geq 2$ ,

$$\mathbb{E}|\chi_{(i,a)}|^p = p \int_0^\infty t^{p-1} \mathbb{P}(|\chi_{(i,a)}| \geq t) dt \leq ep \int_0^\infty t^{p-1} e^{-t} dt = 2e \frac{p!}{2}. \quad (64)$$

For the last equality we used that the appearing integral is equal to  $\Gamma(p) = (p-1)!$ , where  $\Gamma$  denotes the well-known Gamma function. Finally, by combining (63) with (64) and recalling (59) we observe that

$$\mathbb{E}[\widehat{\mathbf{Y}}^{(i,a)}]^p \preceq \frac{p!}{2} \left(\frac{\max_{\tau \in [N_t]} |S_{[\beta]}^\tau|}{N_T N_t}\right)^{p-2} [\sqrt{2e} \mathbf{Y}^{(i,a),(i,a)}]^2.$$



With the matrices  $\sqrt{2e}\mathbf{Y}^{(i,a),(i,a)}$  in place of the matrices  $\mathbf{A}_k$ , the quantity (61) can be calculated as

$$\sigma^2 = \left\| \sum_{i=1}^{N_T} \sum_{a=1}^{N_t} [\sqrt{2e}\mathbf{Y}^{(i,a),(i,a)}]^2 \right\|_{2 \rightarrow 2} = 2e \left\| \sum_{i=1}^{N_T} \sum_{a=1}^{N_t} \mathbf{Y}^{(i,a),(i,a)} \mathbf{Y}^{(i,a),(i,a)} \right\|_{2 \rightarrow 2} = 2e \frac{\max_{\tau \in [N_t]} |S_{[\beta]}^\tau|}{N_T N_t},$$

where we used that

$$\left[ \sum_{i=1}^{N_T} \sum_{a=1}^{N_t} \mathbf{Y}^{(i,a),(i,a)} \mathbf{Y}^{(i,a),(i,a)} \right]_{\Theta, \Theta'} = \delta_{\Theta, \Theta'} \frac{|S_{[\beta]}^\tau|}{N_T N_t}.$$

(The proof is analog to the identities in (81), see the appendix.) A direct application of Theorem 23, where the first bound in (62) applies, yields

$$\mathbb{P}(\|\mathbf{Y}_= - \mathbf{Id}\|_{2 \rightarrow 2} \geq \delta/2) \leq |S_{[\beta]}| \exp \left( -\frac{(\delta/2)^2 N_T N_t}{4 \cdot 2e \max_{\tau \in [N_t]} |S_{[\beta]}^\tau|} \right) \leq |S_{[\beta]}| \exp \left( -\frac{\delta^2 N_T N_t}{32e |S_{[\beta]}|} \right). \quad (65)$$

### A tail bound for $\|\mathbf{Y}_\neq\|_{2 \rightarrow 2}$

In order to derive a tail estimate for the quantity  $\|\mathbf{Y}_\neq\|_{2 \rightarrow 2}$ , we examine the moments  $\mathbb{E}\|\mathbf{Y}_\neq\|_{2 \rightarrow 2}^{2m}$ ,  $m \in \mathbb{N}$ . We use decoupling (see, e.g., [14, Thm. 8.11]) in order to obtain

$$\mathbb{E}\|\mathbf{Y}_\neq\|_{2 \rightarrow 2}^{2m} = \mathbb{E} \left\| \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \overline{s_{(i,a)} s_{(j,b)}} \mathbf{Y}^{(i,a),(j,b)} \right\|_{2 \rightarrow 2}^{2m} \leq 4^{2m} \mathbb{E} \left\| \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \xi_{(i,a)} \xi'_{(j,b)} \mathbf{Y}^{(i,a),(j,b)} \right\|_{2 \rightarrow 2}^{2m}, \quad (66)$$

where the  $\xi_{(i,a)}, \xi'_{(j,b)}$ ,  $(i,a), (j,b) \in [N_T] \times [N_t]$ , are independent standard complex Gaussian random variables. The term on the right-hand side can be estimated by means of the following Khintchine-type inequality for a homogeneous matrix-valued chaos of order two taken from [25]<sup>2</sup>. Although the original result [25, Thm. 6.22] is formulated for independent Rademacher random variables, a careful observation of the proof reveals that it is also true for standard complex Gaussians.

**Lemma 24.** *Let  $\mathbf{B}_{k,\ell}$ ,  $k, \ell = 1, \dots, M$ , be complex matrices of the same dimension and let  $\xi_k, \xi'_\ell$ ,  $k, \ell \in [M]$ , independent standard Gaussian random variables. Then, for  $m \in \mathbb{N}$ ,*

$$\begin{aligned} & \mathbb{E} \left\| \sum_{k,\ell=1}^M \xi_k \xi'_\ell \mathbf{B}_{k,\ell} \right\|_{S_{2m}}^{2m} \\ & \leq \left( \frac{(2m)!}{2^m m!} \right)^2 \max \left\{ \left\| \left( \sum_{k,\ell=1}^M \mathbf{B}_{k,\ell} \mathbf{B}_{k,\ell}^* \right)^{1/2} \right\|_{S_{2m}}^{2m}, \left\| \left( \sum_{k,\ell=1}^M \mathbf{B}_{k,\ell}^* \mathbf{B}_{k,\ell} \right)^{1/2} \right\|_{S_{2m}}^{2m}, \|\mathbf{F}\|_{S_{2m}}^{2m}, \|\tilde{\mathbf{F}}\|_{S_{2m}}^{2m} \right\}, \end{aligned} \quad (67)$$

where  $\mathbf{F}, \tilde{\mathbf{F}}$  are the block matrices  $\mathbf{F} = (\mathbf{B}_{k,\ell})_{k,\ell=1}^M$  and  $\tilde{\mathbf{F}} = (\mathbf{B}_{k,\ell}^*)_{k,\ell=1}^M$ , respectively.

In our setting, the matrices  $\mathbf{B}_{k,\ell}$  are given by the matrices  $\mathbf{Y}^{(i,a),(j,b)}$ . Therefore,  $\mathbf{F}$  and  $\tilde{\mathbf{F}}$  appearing in the assertion of Lemma 24 are block matrices consisting of  $|S_{[\beta]}| \times |S_{[\beta]}|$  blocks. In the following we calculate the Schatten  $2m$ -norms of the matrix  $\mathbf{F}$ . To this end we compute the eigenvalues of  $\mathbf{F}^* \mathbf{F}$ . Like the matrix  $\mathbf{F}$  itself also the product  $\mathbf{F}^* \mathbf{F}$  consists of  $|S_{[\beta]}|$ -dimensional blocks. A straightforward calculation using the definition of the matrices reveals (see the end of Appendix C for the details) that the block at position  $(i,a), (j,b)$  of this matrix product is entrywise given by

$$[\mathbf{F}^* \mathbf{F}]_{\Theta, \Theta'}^{(i,a),(j,b)} = \delta_{\tau'+b, \tau+a}^{-1} N_T^{-1} N_t^{-2} |S_{[\beta]}| e^{i2\pi \cdot \frac{f'-f}{N_t}(\tau+a-1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]}.$$

<sup>2</sup> Note that due to the comment after the formulation of the original result in [25] the extra factor  $2^{1/2n}$  appearing right-hand side of the assertion in [25, Thm. 6.22] can be removed. In the original version of [25, Thm. 6.22], the quantity  $\|\tilde{\mathbf{F}}\|_{S_{2m}}^{2m}$  was missing; this has been corrected in a new version which can be obtained from the personal website of HR.

Due to the appearance of the factor  $\delta_{\tau'+b, \tau+a}^{\sim N_t}$  the matrix  $\mathbf{F}^* \mathbf{F}$  can be transformed into a block diagonal matrix by merely rearranging the ordering of the rows/columns. Hence, we assume without loss of generality that  $\mathbf{F}^* \mathbf{F}$  is block diagonal, with the blocks  $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \dots, \mathbf{\Lambda}_{N_t}$  on the diagonal, where each block  $\mathbf{\Lambda}_\kappa$  is a rank one matrix of the form

$$\mathbf{\Lambda}_\kappa = \frac{|S_{[\beta]}|}{N_T N_t^2} \mathbf{w}_\kappa \mathbf{w}_\kappa^*,$$

and where for each  $\kappa \in [N_t]$  the vector  $\mathbf{w}_\kappa \in \mathbb{C}^{N_T |S_{[\beta]}|}$  is entrywise given by

$$[\mathbf{w}_\kappa]_{i, \Theta} = e^{i2\pi \cdot \frac{f}{N_t} (\kappa-1)} e^{i2\pi \cdot d_T \Delta_{\beta} \beta(i-1)}, \quad i \in [N_T], \Theta \in S_{[\beta]}.$$

Hence,  $\|\mathbf{w}_\kappa\|_2^2 = N_T |S_{[\beta]}|$  and, therefore,  $\mathbf{w}_\kappa \mathbf{w}_\kappa^*$  has exactly one nonzero eigenvalue, equal to  $N_T |S_{[\beta]}|$ . This implies that also the block  $\mathbf{\Lambda}_\kappa$  has exactly one nonzero eigenvalue which is equal to  $|S_{[\beta]}|^2 / N_t^2$  and, thus,  $\|\mathbf{\Lambda}_\kappa\|_{S_m} = |S_{[\beta]}|^2 / N_t^2$ . With this we can calculate,

$$\|\mathbf{F}\|_{S_{2m}}^{2m} = \|\mathbf{F}^* \mathbf{F}\|_{S_m}^m = \sum_{\kappa \in [N_t]} \|\mathbf{\Lambda}_\kappa\|_{S_m}^m = N_t \frac{|S_{[\beta]}|^{2m}}{N_t^{2m}} = |S_{[\beta]}| \frac{|S_{[\beta]}|^{2m-1}}{N_t^{2m-1}} \leq |S_{[\beta]}| \frac{|S_{[\beta]}|^m}{N_t^m}, \quad (68)$$

where we used that, without loss of generality,  $|S_{[\beta]}| \leq N_t$ . By using the same arguments for the matrix  $\tilde{\mathbf{F}}$  it is straightforward to verify that also

$$\|\tilde{\mathbf{F}}\|_{S_{2m}}^{2m} \leq |S_{[\beta]}| \frac{|S_{[\beta]}|^m}{N_t^m} \quad (69)$$

holds true.

We proceed with estimating the first two quantities in the maximum expression in (67). As we calculate in Appendix C (see (81)),

$$\sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \mathbf{Y}^{(i,a),(j,b)} [\mathbf{Y}^{(i,a),(j,b)}]^* = \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} [\mathbf{Y}^{(i,a),(j,b)}]^* \mathbf{Y}^{(i,a),(j,b)} = \frac{|S_{[\beta]}|}{N_t} \mathbf{Id}.$$

Therefore, the corresponding Schatten  $2m$ -norm in (67) can be calculated as

$$\left\| \left[ \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \mathbf{Y}^{(i,a),(j,b)} [\mathbf{Y}^{(i,a),(j,b)}]^* \right]^{1/2} \right\|_{S_{2m}}^{2m} = \left\| \left[ \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} [\mathbf{Y}^{(i,a),(j,b)}]^* \mathbf{Y}^{(i,a),(j,b)} \right]^{1/2} \right\|_{S_{2m}}^{2m} = |S_{[\beta]}| \frac{|S_{[\beta]}|^m}{N_t^m}.$$

Now a direct application of the Khintchine-type inequality (see Lemma 24 above) to the moments in (66), using (68), (69), and the latter equalities implies

$$\mathbb{E} \|\mathbf{Y}_\neq\|_{2 \rightarrow 2}^{2m} \leq 4^{2m} \left( \frac{(2m)!}{2^m m!} \right)^2 |S_{[\beta]}| \frac{|S_{[\beta]}|^m}{N_t^m} = \left( \frac{(2m)!}{2^m m!} \right)^2 |S_{[\beta]}| \left( \frac{4\sqrt{|S_{[\beta]}|}}{\sqrt{N_t}} \right)^{2m}. \quad (70)$$

Using Stirling's formula, one can easily verify that

$$\frac{(2m)!}{2^m m!} \leq \sqrt{2} (2/e)^m m^m,$$

Therefore, the moments in (70) can further be estimated as

$$\mathbb{E} \|\mathbf{Y}_\neq\|_{2 \rightarrow 2}^{2m} \leq 2 |S_{[\beta]}| \left( \frac{8\sqrt{|S_{[\beta]}|}}{e\sqrt{N_t}} \right)^{2m} m^{2m}, \quad (71)$$

which holds true for  $m \in \mathbb{N} \cup \{0\}$ . Hölder's inequality implies

$$\mathbb{E} |Z|^{2m+2\theta} = \mathbb{E} [|Z|^{(1-\theta)2m} |Z|^{\theta(2m+2)}] \leq (\mathbb{E} |Z|^{2m})^{1-\theta} (\mathbb{E} |Z|^{2m+2})^\theta,$$

for any random variable  $Z$ , and each  $\theta \in [0, 1]$ . Combining this with (71) gives

$$\mathbb{E}\|\mathbf{Y}_{\neq}\|_{2 \rightarrow 2}^{2m+2\theta} \leq 2|S_{[\beta]}| \left( \frac{8\sqrt{|S_{[\beta]}|}}{e\sqrt{N_t}} \right)^{2m+2\theta} m^{2m(1-\theta)} (m+1)^{(2m+2)\theta}.$$

For  $m \in \mathbb{N}$  we estimate the last terms as

$$m^{2m(1-\theta)} (m+1)^{(2m+2)\theta} = (m^{1-\theta} (m+1)^\theta)^{2m+2\theta} \left( \frac{m+1}{m} \right)^{2\theta(1-\theta)} \leq \sqrt{2} (m+\theta)^{2m+2\theta}, \quad (72)$$

where we used the inequality between the (weighted) arithmetic and geometric mean and, furthermore, used the fact that  $(m+1)/m \leq 2$  and  $\theta(1-\theta) \leq 1/4$ . Since

$$\inf\{x^x : x \geq 0\} = e^{-1/e} \geq 3/5,$$

we can replace  $\sqrt{2}$  on the right-hand side in (72) by 3 leaving us with an inequality which is now also valid for  $m = 0$ . Altogether it holds for all real  $p = 2m + 2\theta > 0$ ,

$$(\mathbb{E}\|\mathbf{Y}_{\neq}\|_{2 \rightarrow 2}^p)^{1/p} \leq \frac{4\sqrt{|S_{[\beta]}|}}{e\sqrt{N_t}} (6|S_{[\beta]}|)^{1/p} p.$$

In order to obtain a tail estimate for the random variable  $\|\mathbf{Y}_{\neq}\|_{2 \rightarrow 2}$  it only remains to apply the following lemma which is a direct implication of Markov's inequality. For a proof see, e.g., [25, 14].

**Lemma 25.** *Let  $Z$  be a random variable and suppose there exist constants  $\alpha, \beta, \gamma, p_0 > 0$  such that*

$$(\mathbb{E}|Z|^p)^{1/p} \leq \alpha\beta^{1/p} p^{1/\gamma},$$

*for all  $p \geq p_0$ . Then it holds for all  $u \geq p_0^{1/\gamma}$ ,*

$$\mathbb{P}(|Z| \geq e^{1/\gamma} \alpha u) \leq \beta e^{-u^\gamma/\gamma}.$$

By applying Lemma 25 we obtain

$$\mathbb{P}(\|\mathbf{Y}_{\neq}\|_{2 \rightarrow 2} \geq \delta/2) \leq 6|S_{[\beta]}| \exp\left(-\frac{\delta\sqrt{N_t}}{8\sqrt{|S_{[\beta]}|}}\right). \quad (73)$$

## Conclusion

Finally, due to the definition of the matrices  $\mathbf{Y}_{\neq}, \mathbf{Y}_{=}$  in (57), an application of the union bound and plugging in the tail bounds (65), (73) (noting that (65) is dominated by (73)) implies the assertion of Proposition 13.  $\square$

## 4 Numerical experiments

We conduct numerical experiments in order to test the effective empirical probabilities of perfect reconstruction of support sets using the LASSO functional, as considered in Theorem 7. Due to runtime limitations we only perform simulations for the Doppler-free scenario. As pointed out in Section 1.6, this case is contained in our analysis as a special case. This means in particular that the same influence of the parameter  $\eta$  on the recovery properties can be expected. This is indeed supported by the results of the numerical experiments described next.

We consider the case where  $N_T = N_R = 8$  and  $N_t = 64$ , yielding a problem dimension of  $N = N_T N_R N_t = 4096$  and a total number of  $m = N_R N_t = 512$  measurements. We run simulations with varying choices for  $(|S|, \eta)$ , where  $|S|$  is the number of targets and  $\eta$  defines the actual balancedness of the considered support sets (see Definition 6). For each such pair  $(|S|, \eta)$  we try to reconstruct 1000 independent random target scenes  $\mathbf{x}$  from measurements  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$  by minimizing the LASSO functional (19) with  $\lambda = 2\sigma\sqrt{2N_T N_R N_t \log(N)}$ , and where, in each run,

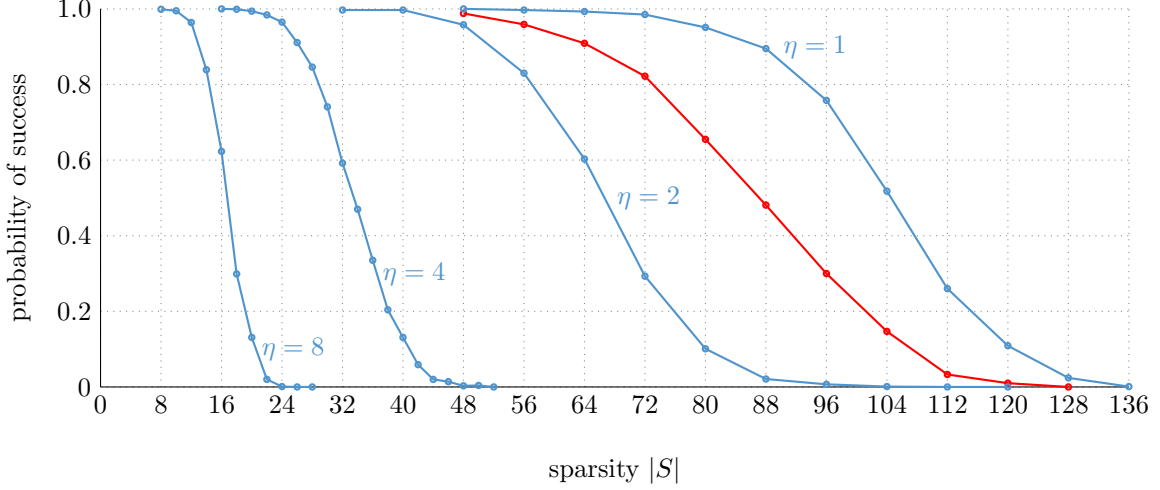


Figure 2: Probability of perfect reconstruction of support sets for varying sparsity levels  $|S|$ . Blue plots: Randomly chosen support sets having constant balancedness parameter  $\eta$  as indicated. Red plot: Randomly chosen support sets with no restriction on the balancedness parameter.

- the signal vectors  $\mathbf{s}_i$  (generating the measurement matrix  $\mathbf{A}$ ) and also the noise vector  $\mathbf{n}$  are drawn randomly from standard complex Gaussian distribution, i.e., we choose  $\sigma^2 = 1$  as the variance of the noise,
- the support set  $S$  of  $\mathbf{x}$  is drawn uniformly at random from all possible support sets having balancedness parameter  $\eta$ ,
- the phases of the nonzero entries of  $\mathbf{x}$  are drawn uniformly at random from  $[0, 2\pi)$ ,
- and the amplitudes  $|x_\theta|$ ,  $\theta \in S$ , are set to the threshold level defined in (26).

We use the Matlab library TFOCS [2] for the minimization of the LASSO functional (19). Since TFOCS uses an iterative solver and, hence, in general only approximations to the desired minimizer can be computed, and we consider a target scene  $\mathbf{x}$  to be successfully recovered if  $\|\mathbf{x} - \mathbf{x}^\# \|_\infty \leq \mathcal{T}$  for an appropriate thresholding  $\mathcal{T} > 0$ .

In Figure 2 we depict the results of our simulations. It becomes evident that small values of the parameter  $\eta$  have a positive impact on the probability of perfect reconstruction of the support set  $S$ . Moreover, as predicted by formula (25), the dependence of the number of maximal reconstructable targets  $|S|$  (for a fixed number of measurements  $N_R N_t$ ) on the parameter  $\eta$  is linear.

## Acknowledgment

The authors acknowledge funding from the European Research Council through the Starting Grant StG 258926.

## Appendix

### A Orthogonality of the matrices $\mathcal{X}_\Theta$

Due to the definition in (31), for  $(i, j) \in [N_R] \times [N_T]$ , the  $(i, j)$ th  $N_t \times N_t$  block of  $\mathcal{X}_\Theta$  is given by

$$\mathcal{X}_\Theta^{[i,j]} = e^{i2\pi \cdot d_R \beta \Delta_\beta(i-1)} e^{i2\pi \cdot d_T \beta \Delta_\beta(j-1)} \mathbf{M}_f \mathbf{T}_\tau.$$

**Lemma 26.** *The set of matrices  $\left\{ \frac{1}{\sqrt{N_t N_R N_T}} \mathcal{X}_\Theta, \Theta \in \mathcal{G} \right\}$  forms an orthonormal basis.*

*Proof.* We calculate the inner product  $\langle \mathbf{X}_{\Theta'}, \mathbf{X}_{\Theta} \rangle = \text{Tr}(\mathbf{X}_{\Theta'}^* \mathbf{X}_{\Theta})$ . To this end we calculate the  $(j', j)$ th block of the product  $\mathbf{X}_{\Theta'}^* \mathbf{X}_{\Theta}$  (recall,  $\mathbf{X}_{\Theta'}^* \mathbf{X}_{\Theta}$  is a  $N_T \times N_T$  block matrix consisting of  $N_t \times N_t$  blocks),

$$\begin{aligned}
[\mathbf{X}_{\Theta'}^* \mathbf{X}_{\Theta}]^{[j', j]} &= \sum_{k=1}^{N_R} [\mathbf{X}_{\Theta'}^*]^{[j', k]} \mathbf{X}_{\Theta}^{[k, j]} = \sum_{k=1}^{N_R} [\mathbf{X}_{\Theta'}^{[k, j']}]^* \mathbf{X}_{\Theta}^{[k, j]} \\
&= \sum_{k=1}^{N_R} \overline{e^{i2\pi \cdot d_R \beta' \Delta_{\beta}(k-1)} e^{i2\pi \cdot d_T \beta' \Delta_{\beta}(j'-1)} e^{i2\pi \cdot d_R \beta \Delta_{\beta}(k-1)} e^{i2\pi \cdot d_T \beta \Delta_{\beta}(j-1)} \mathbf{T}_{\tau'}^* \mathbf{M}_{f'}^* \mathbf{M}_f \mathbf{T}_{\tau}} \\
&= e^{-i2\pi \cdot d_T \beta' \Delta_{\beta}(j'-1)} e^{i2\pi \cdot d_T \beta \Delta_{\beta}(j-1)} e^{i2\pi \cdot \frac{(f-f')}{N_t} \tau'} \sum_{k=1}^{N_R} e^{i2\pi \cdot d_R (\beta - \beta') \Delta_{\beta}(k-1)} \mathbf{M}_{f-f'} \mathbf{T}_{\tau-\tau'}
\end{aligned} \tag{74}$$

For the last equality we used that

$$\mathbf{T}_{\tau'}^* \mathbf{M}_{f'}^* \mathbf{M}_f \mathbf{T}_{\tau} = e^{i2\pi \cdot \frac{(f-f')}{N_t} \tau'} \mathbf{M}_{f-f'} \mathbf{T}_{\tau-\tau'},$$

which follows directly from the definitions of the operators  $\mathbf{M}_f$  and  $\mathbf{T}_{\tau}$  (see (6)). Due to (74), the Frobenius inner product between two matrices is given as

$$\begin{aligned}
\langle \mathbf{X}_{\Theta'}, \mathbf{X}_{\Theta} \rangle &= \sum_{j=1}^{N_T} \text{Tr}([\mathbf{X}_{\Theta'}^* \mathbf{X}_{\Theta}]^{[j, j]}) \\
&= \left( \sum_{j=1}^{N_T} e^{i2\pi \cdot d_T (\beta - \beta') \Delta_{\beta}(j-1)} \right) e^{i2\pi \cdot \frac{(f-f')}{N_t} \tau'} \sum_{k=1}^{N_R} e^{i2\pi \cdot d_R (\beta - \beta') \Delta_{\beta}(k-1)} \text{Tr}(\mathbf{M}_{f-f'} \mathbf{T}_{\tau-\tau'}).
\end{aligned}$$

Since  $\mathbf{M}_{f-f'}$  is a diagonal matrix, the trace of the product  $\mathbf{M}_{f-f'} \mathbf{T}_{\tau-\tau'}$  can only be nonzero if at least one of the diagonal entries of the matrix  $\mathbf{T}_{\tau-\tau'}$  is nonzero, i.e., if  $\tau = \tau'$  so that  $\mathbf{T}_{\tau-\tau'} = \mathbf{Id}$ . This means that

$$\text{Tr}(\mathbf{M}_{f-f'} \mathbf{T}_{\tau-\tau'}) = \text{Tr}(\mathbf{M}_{f-f'}) = \sum_{k=1}^{N_t} e^{i2\pi \cdot \frac{f-f'}{N_t} (k-1)} = \begin{cases} N_t & \text{if } f = f', \\ 0 & \text{otherwise.} \end{cases}$$

Recalling the formula for  $\langle \mathbf{X}_{\Theta'}, \mathbf{X}_{\Theta} \rangle$  from above implies that for this inner product to be nonzero it necessarily has to hold that  $\Theta' = \Theta$ . Indeed, this follows from the appearance of the factor  $\sum_{j=1}^{N_T} e^{i2\pi \cdot d_T (\beta - \beta') \Delta_{\beta}(j-1)}$  which (recalling that  $d_T = 1/2$  and  $\Delta_{\beta} = 2/N_T N_R$ , see (1), (4)) is only nonzero (and equal to  $N_T$ ) if  $\beta' = \beta$ . Finally, we can conclude

$$\langle \mathbf{X}_{\Theta'}, \mathbf{X}_{\Theta} \rangle = \delta_{\Theta', \Theta} N_T N_R N_t.$$

The normalization yields the result.  $\square$

## B Proof of Lemma 11

For small  $u$  the first term on the right-hand side of (37) can be obtained by a volumetric argument. To this end, let for  $S \subset \{1, \dots, N\}$ ,  $B_S \subset \mathbb{C}^N$  denote the set of all vectors  $\mathbf{x}$  with  $\|\mathbf{x}\|_2 \leq 1$  and support in  $S$ . Introducing  $\|\mathbf{x}\| := \|\tilde{\mathbf{V}}_{\mathbf{x}}\|_{2 \rightarrow 2}$  we find using (35) and a volumetric estimate, see e.g. [14, Proposition C.3],

$$\mathcal{N}(B_S, \|\cdot\|, u) \leq \mathcal{N}(B_S, \sqrt{s/N_t} \|\cdot\|_2, u) \leq \left(1 + 2 \frac{\sqrt{s/N_t}}{u}\right)^{2s} \leq \left(3 \frac{\sqrt{s/N_t}}{u}\right)^{2s},$$

where for the last inequality we used the assumption  $u \leq \sqrt{s/N_t}$ . Since  $\mathcal{A}$  is the union of all sets  $B_S$  with  $S \subset [N]$ , and there are  $\binom{N}{s} \leq (eN/s)^s$  possible choices for  $S$  it holds

$$\mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \leq (eN/s)^s (3\sqrt{s/N_t}/u)^{2s},$$

which implies the first bound in (37), namely

$$\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \leq 2s \left( \log(eN/s) + \log \left( \frac{2\sqrt{s}}{u\sqrt{N_t}} \right) \right) \lesssim s \log \left( \frac{N}{u^2 N_t} \right).$$

For the second bound from the assertion we exploit the fact that

$$\{\mathbf{x} \in \mathbb{C}^N : \mathbf{x} \text{ } s\text{-sparse}, \|\mathbf{x}\|_2 \leq 1\} \subset \sqrt{2s} \operatorname{conv} \bigcup_{\theta \in \mathcal{G}} \{\mathbf{e}_\theta, \imath \mathbf{e}_\theta, -\mathbf{e}_\theta, -\imath \mathbf{e}_\theta\} =: \tilde{D}_s$$

and, hence, the set  $\mathcal{A}$  from (33) is contained in the set  $\{\tilde{\mathbf{V}}_{\mathbf{x}} : \mathbf{x} \in \tilde{D}_s\}$ . The following is a version of *Maurey's lemma*. For a proof see, e.g., [20].

**Lemma 27.** *There exists an absolute constant  $c$  for which the following holds. Let  $X$  be a normed space, consider a finite set  $\mathcal{U} \subset X$  of cardinality  $N$ , and assume that for every  $L \in \mathbb{N}$  and  $(\mathbf{u}_1, \dots, \mathbf{u}_L) \in \mathcal{U}^L$ ,  $\mathbb{E} \|\sum_{j=1}^L \epsilon_j \mathbf{u}_j\|_X \leq A\sqrt{L}$ , where  $(\epsilon_1, \dots, \epsilon_L)$  denotes a Rademacher vector. Then for every  $u > 0$ ,*

$$\log \mathcal{N}(\operatorname{conv}(\mathcal{U}), \|\cdot\|_X, u) \leq c(A/u)^2 \log(N).$$

In order to apply Lemma 27, we need to estimate the quantity  $\mathbb{E}_{\epsilon} \|\sum_{k=1}^L \epsilon_k \tilde{\mathbf{V}}_{\mathbf{u}_k}\|_{2 \rightarrow 2}$ , where  $(\mathbf{u}_1, \dots, \mathbf{u}_L)$  is a sequence of extreme points in  $\tilde{D}_s$  and  $\epsilon = (\epsilon_1, \dots, \epsilon_L)$  is a Rademacher vector. The noncommutative Khintchine inequality [3, 25] – originally due to Lust-Piquard [23, 24] – or more modern estimates based on moment generating function bounds [30], see also [14, Problem 8.6(d)],

$$\mathbb{E}_{\epsilon} \left\| \sum_{k=1}^L \epsilon_k \tilde{\mathbf{V}}_{\mathbf{u}_k} \right\|_{2 \rightarrow 2} \lesssim \sqrt{\log(N_{\max})} \max \left\{ \left\| \sum_{k=1}^L \tilde{\mathbf{V}}_{\mathbf{u}_k} \tilde{\mathbf{V}}_{\mathbf{u}_k}^* \right\|_{2 \rightarrow 2}, \left\| \sum_{k=1}^L \tilde{\mathbf{V}}_{\mathbf{u}_k}^* \tilde{\mathbf{V}}_{\mathbf{u}_k} \right\|_{2 \rightarrow 2} \right\}^{1/2}, \quad (75)$$

where  $N_{\max}$  stands for the maximum of the dimensions of the matrices  $\tilde{\mathbf{V}}_{\mathbf{u}_k} \tilde{\mathbf{V}}_{\mathbf{u}_k}^*$  and  $\tilde{\mathbf{V}}_{\mathbf{u}_k}^* \tilde{\mathbf{V}}_{\mathbf{u}_k}$ , and can be estimated by  $\max\{N_R N_t, N_T N_t\} \leq N$ . Using the estimate (35) for  $\|\tilde{\mathbf{V}}_{\mathbf{u}_k}\|_{2 \rightarrow 2}$ ,

$$\|\tilde{\mathbf{V}}_{\mathbf{u}_k} \tilde{\mathbf{V}}_{\mathbf{u}_k}^*\|_{2 \rightarrow 2} = \|\tilde{\mathbf{V}}_{\mathbf{u}_k}^* \tilde{\mathbf{V}}_{\mathbf{u}_k}\|_{2 \rightarrow 2} = \|\tilde{\mathbf{V}}_{\mathbf{u}_k}\|_{2 \rightarrow 2}^2 \leq \frac{1}{N_t} \|\mathbf{u}_k\|_1^2 = \frac{2s}{N_t}.$$

An application of the triangle inequality yields, using the Khintchine inequality (75),

$$\mathbb{E}_{\epsilon} \left\| \sum_{k=1}^L \epsilon_k \tilde{\mathbf{V}}_{\mathbf{u}_k} \right\|_{2 \rightarrow 2} \lesssim \sqrt{\log(N)} \sqrt{\frac{2s}{N_t}} \sqrt{L}$$

Finally, we can apply Lemma 27 yielding

$$\log \mathcal{N}(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \lesssim \frac{s}{u^2 N_t} \log^2(N).$$

This establishes the second bound in (37).  $\square$

## C Basic calculations for Proposition 13

The proof of Proposition 13 is based on the fact that

$$\tilde{\mathbf{A}}_{S[\beta]}^* \tilde{\mathbf{A}}_{S[\beta]} = \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \overline{s_{(i,a)} s_{(j,b)}} \mathbf{Y}^{(i,a),(j,b)}, \quad (76)$$

where we write  $s_{(i,a)}$  for the  $a$ -th entry of the signal vector  $\mathbf{s}_i$  and where for  $\theta, \theta' \in S[\beta]$  the corresponding entry of a given matrix  $\mathbf{Y}^{(i,a),(j,b)}$  is given by

$$[\mathbf{Y}^{(i,a),(j,b)}]_{\theta, \theta'} = \delta_{a-b, \tau' - \tau}^{N_t} (N_T N_t)^{-1} e^{i2\pi \cdot \frac{f' - f}{N_t} (\tau + a - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]}, \quad (77)$$

To see this we recall that, according to (22), the inner products  $\langle \mathbf{A}_\Theta, \mathbf{A}_{\Theta'} \rangle$  satisfy

$$\begin{aligned} [\tilde{\mathbf{A}}_{S[\beta]}^* \tilde{\mathbf{A}}_{S[\beta]}]_{\Theta, \Theta'} &= \langle \tilde{\mathbf{A}}_\Theta, \tilde{\mathbf{A}}_{\Theta'} \rangle = (N_T N_R N_t)^{-1} \langle \mathbf{A}_\Theta, \mathbf{A}_{\Theta'} \rangle \\ &= (N_T N_t)^{-1} \sum_{i,j=1}^{N_T} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]} \langle \mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i, \mathbf{M}_{f'} \mathbf{T}_{\tau'} \mathbf{s}_j \rangle, \end{aligned}$$

where we used that both  $\Theta, \Theta' \in S[\beta]$ . Recalling the definitions of the operators  $\mathbf{M}_f, \mathbf{T}_\tau$  (see (6)) one obtains for the latter inner product,

$$\begin{aligned} \langle \mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i, \mathbf{M}_{f'} \mathbf{T}_{\tau'} \mathbf{s}_j \rangle &= \sum_{k=1}^{N_t} \overline{[\mathbf{M}_f \mathbf{T}_\tau \mathbf{s}_i]_k} [\mathbf{M}_{f'} \mathbf{T}_{\tau'} \mathbf{s}_j]_k = \sum_{k=1}^{N_t} e^{i2\pi \cdot \frac{f' - f}{N_t} (k-1)} \overline{[\mathbf{s}_i]_{k-\tau}} [\mathbf{s}_j]_{k-\tau'} \\ &= \sum_{a,b=1}^{N_t} \delta_{a-b, \tau' - \tau}^{\sim N_t} e^{i2\pi \cdot \frac{f' - f}{N_t} (a + \tau - 1)} \overline{[\mathbf{s}_i]_a} [\mathbf{s}_j]_b. \end{aligned}$$

By combining the identities from above one finds

$$[\tilde{\mathbf{A}}_{S[\beta]}^* \tilde{\mathbf{A}}_{S[\beta]}]_{\Theta, \Theta'} = \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \overline{s_{(i,a)}} s_{(j,b)} \underbrace{\delta_{a-b, \tau' - \tau}^{\sim N_t} (N_T N_t)^{-1} e^{i2\pi \cdot \frac{f' - f}{N_t} (\tau + a - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]}}_{=[\mathbf{Y}^{(i,a), (j,b)}]_{\Theta, \Theta'}, \text{ see (77)}},$$

which shows (76).

The matrices  $\mathbf{Y}^{(i,a), (j,b)}$  allow for a simple formula for their adjoints, namely

$$[\mathbf{Y}^{(i,a), (j,b)}]^* = \mathbf{Y}^{(j,b), (i,a)}. \quad (78)$$

### The product $\mathbf{F}^* \mathbf{F}$

The matrix  $\mathbf{F}$  consists of the blocks  $\mathbf{Y}^{(i,a), (j,b)}$  given by (77). Therefore,  $\mathbf{F}$  is self-adjoint so that  $\mathbf{F}^* \mathbf{F} = \mathbf{F}^2$ . Like  $\mathbf{F}$  also the product  $\mathbf{F}^2$  consists of blocks and the block at the  $(i, a)$ -th (block) row and the  $(j, b)$ -th (block) column is given by

$$[\mathbf{F}^2]^{(i,a), (j,b)} = \sum_{r=1}^{N_T} \sum_{q=1}^{N_t} \mathbf{Y}^{(i,a), (r,q)} \mathbf{Y}^{(r,q), (j,b)}. \quad (79)$$

Recalling (77), the appearing summands  $\mathbf{Y}^{(i,a), (r,q)} \mathbf{Y}^{(r,q), (j,b)}$  are given entrywise by

$$\begin{aligned} [\mathbf{Y}^{(i,a), (r,q)} \mathbf{Y}^{(r,q), (j,b)}]_{\Theta, \Theta'} &= \sum_{\tilde{\Theta} \in S[\beta]} \mathbf{Y}_{\Theta, \tilde{\Theta}}^{(i,a), (r,q)} \mathbf{Y}_{\tilde{\Theta}, \Theta'}^{(r,q), (j,b)} \\ &= (N_T N_t)^{-2} \sum_{\tilde{\Theta} \in S[\beta]} \delta_{a-q, \tilde{\tau} - \tau}^{\sim N_t} \delta_{q-b, \tau' - \tilde{\tau}}^{\sim N_t} e^{i2\pi \cdot \frac{\tilde{f} - f}{N_t} (\tau + a - 1)} e^{i2\pi \cdot \frac{f' - \tilde{f}}{N_t} (\tilde{\tau} + q - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]} \\ &= (N_T N_t)^{-2} \delta_{\tau' + b, \tau + a}^{\sim N_t} |S_{[\beta]}^{\tau + a - q}| e^{i2\pi \cdot \frac{f' - f}{N_t} (a + \tau - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]}. \end{aligned} \quad (80)$$

Combining this with (79) yields

$$\begin{aligned} [\mathbf{F} \mathbf{F}]_{\Theta, \Theta'}^{(i,a), (j,b)} &= \sum_{r=1}^{N_T} \sum_{q=1}^{N_t} \delta_{\tau' + b, \tau + a}^{\sim N_t} (N_T N_t)^{-2} |S_{[\beta]}^{\tau + a - q}| e^{i2\pi \cdot \frac{f' - f}{N_t} (a + \tau - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]} \\ &= \delta_{\tau' + b, \tau + a}^{\sim N_t} N_T^{-1} N_t^{-2} |S_{[\beta]}| e^{i2\pi \cdot \frac{f' - f}{N_t} (a + \tau - 1)} e^{i2\pi \cdot d_T \Delta_\beta [\beta'(j-1) - \beta(i-1)]}. \end{aligned}$$



### Properties of the matrices $\mathbf{Y}^{(i,a),(j,b)}$

The proof of Proposition 13 uses the identities

$$\sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} [\mathbf{Y}^{(i,a),(j,b)}]^* \mathbf{Y}^{(i,a),(j,b)} = \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \mathbf{Y}^{(i,a),(j,b)} [\mathbf{Y}^{(i,a),(j,b)}]^* = \frac{|S_{[\beta]}|}{N_t} \mathbf{Id}. \quad (81)$$

Due to (78), and by plugging in the identity we used in the second step of (80), the second sum is given entrywise by

$$\begin{aligned} \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} [\mathbf{Y}^{(i,a),(j,b)} \mathbf{Y}^{(j,b),(i,a)}]_{\Theta, \Theta'} &= \sum_{i,j=1}^{N_T} \sum_{a,b=1}^{N_t} \delta_{\tau', \tau} \frac{|S_{[\beta]}^{\tau+a-b}|}{(N_T N_t)^2} e^{i2\pi \cdot \frac{f'-f}{N_t}(\tau+a-1)} e^{i2\pi \cdot d_T \Delta_\beta(\beta'-\beta)(i-1)} \\ &= \delta_{\tau', \tau} \frac{N_T |S_{[\beta]}|}{(N_T N_t)^2} e^{i2\pi \cdot \frac{f'-f}{N_t} \tau} \sum_{a=1}^{N_t} e^{i2\pi \cdot \frac{f'-f}{N_t}(a-1)} \sum_{i=1}^{N_T} e^{i2\pi \cdot d_T \Delta_\beta(\beta'-\beta)(i-1)} = \delta_{\Theta, \Theta'} \frac{|S_{[\beta]}|}{N_t}, \end{aligned}$$

which establishes the second equality in (81). The first equality follows due to symmetry.

## D Basics from probability theory

A complex-valued random variable  $\xi$  is standard complex Gaussian iff it has (complex) density  $\frac{1}{\pi} e^{-|\xi|^2}$ , or, equivalently,  $\xi$  can be written as  $\xi = x + iy$ , where  $x, y \sim N(0, 1/2)$  are independent standard Gaussian random variables. More generally, a mean-zero complex Gaussian random variable with variance  $\sigma^2$  is of the form  $\sigma\xi$ , where  $\xi$  is a standard complex Gaussian.

**Lemma 28.** *For a standard complex Gaussian random variable  $\xi$  it holds*

$$\mathbb{P}(|\xi| \geq t) \leq e^{-t^2}.$$

For a standard complex Gaussian random vector  $\boldsymbol{\xi}$  (having independent, standard complex Gaussian entries) and a (deterministic) complex vector  $\mathbf{a}$  of the same dimension, the random variable  $z := \langle \mathbf{a}, \boldsymbol{\xi} \rangle$  is mean-zero complex Gaussian with variance  $\|\mathbf{a}\|_2^2$ . This implies the next statement.

**Lemma 29.** *For a standard complex Gaussian random vector  $\boldsymbol{\xi}$  and a complex vector  $\mathbf{a}$  of the same dimension it holds*

$$\mathbb{P}(|\langle \mathbf{a}, \boldsymbol{\xi} \rangle| \geq t) = \mathbb{P}(|\xi| \geq t/\|\mathbf{a}\|_2) \leq e^{-t^2/\|\mathbf{a}\|_2^2}.$$

For a  $2n$ -dimensional standard Gaussian random vector  $\mathbf{g}$  we have, due to [14, (8.89)],

$$\mathbb{P}(\|\mathbf{g}\|_2 \geq \sqrt{2n} + t) \leq e^{-t^2/2}.$$

Since an  $n$ -dimensional standard complex Gaussian random vector  $\boldsymbol{\xi}$  can be considered as a (real-valued)  $2n$ -dimensional standard Gaussian random vector  $\mathbf{g}$  with independent entries from  $\mathcal{N}(0, 1/2)$ , we have the following lemma.

**Lemma 30.** *For an  $n$ -dimensional standard complex Gaussian random vector  $\boldsymbol{\xi}$  it holds*

$$\mathbb{P}(\|\boldsymbol{\xi}\|_2 \geq \sqrt{n} + t) = \mathbb{P}(\|2^{-1/2} \mathbf{g}\|_2 \geq \sqrt{n} + t) = \mathbb{P}(\|\mathbf{g}\|_2 \geq \sqrt{2n} + \sqrt{2}t) \leq e^{-t^2}.$$

Finally, the following lemma states a Hoeffding-type inequality for Steinhaus sequences. Recall that a Steinhaus sequence is a sequence of independent random variables which are all distributed uniformly on the complex unit circle  $\{z \in \mathbb{C} : |z| = 1\}$ .

**Lemma 31** ([14, Cor. 8.10]). *Let  $\mathbf{a} \in \mathbb{C}^L$  and  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_L)$  be a Steinhaus sequence. Then*

$$\mathbb{P}(|\langle \mathbf{a}, \boldsymbol{\epsilon} \rangle| \geq u\|\mathbf{a}\|_2) \leq 2e^{-u^2/2}.$$

## References

- [1] BARANIUK, R. G., AND STEEGHS, P. Compressive radar imaging. In *Proc. IEEE Radar Conf. 2007* (April 2007), pp. 128–133.
- [2] BECKER, S., CANDÈS, E. J., AND GRANT, M. Templates for convex cone problems with applications to sparse signal recovery. *Math. Program. Comput.* 3, 3 (2011), 165–218.
- [3] BUCHHOLZ, A. Operator Khintchine inequality in non-commutative probability. *Math. Ann.* 319 (2001), 1–16.
- [4] CAI, T., AND ZHANG, A. Sparse Representation of a Polytope and Recovery of Sparse Signals and Low-Rank Matrices,. *IEEE Trans. Inform. Theory* 60, 1 (Jan, 2014), 122–132.
- [5] CANDÈS, E. J. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Paris S’er. I Math.* 346 (2008), 589–592.
- [6] CANDÈS, E. J., AND PLAN, Y. Near-ideal model selection by  $\ell_1$  minimization. *Ann. Statist.* 37, 5A (2009), 2145–2177.
- [7] CANDÈS, E. J., ROMBERG, J. K., AND TAO, T. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* 59, 8 (2006), 1207–1223.
- [8] CHEN, C.-Y., AND VAIDYANATHAN, P. Compressed sensing in MIMO radar. In *Signals, Systems and Computers, 2008 42nd Asilomar Conference on* (Oct 2008), pp. 41–44.
- [9] CHI, Y., SCHARF, L., PEZESHKI, A., AND CALDERBANK, A. Sensitivity to basis mismatch in compressed sensing. *Signal Processing, IEEE Transactions on* 59, 5 (2011), 2182–2195.
- [10] COHEN, A., DAHMEN, W., AND DEVORE, R. A. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.* 22, 1 (2009), 211–231.
- [11] DIRKSEN, S. Tail bounds via generic chaining. *Electron. J. Probab.* (to appear).
- [12] ENDER, J. On compressive sensing applied to radar. *Signal Process.* 90, 5 (2010), 1402 – 1414.
- [13] FISHLER, E., HAIMOVICH, A., BLUM, R., CHIZHIK, D., CIMINI, L., AND VALENZUELA, R. MIMO radar: an idea whose time has come. In *Radar Conference, 2004. Proceedings of the IEEE* (April 2004), pp. 71–78.
- [14] FOUCART, S., AND RAUHUT, H. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- [15] FRIEDLANDER, B. On the Relationship between MIMO and SIMO Radars. *Signal Processing, IEEE Transactions on* 57, 1 (Jan 2009), 394–398.
- [16] HECKEL, R., MORGENSTERN, V. I., AND SOLTANOLKOTABI, M. Super-resolution radar. *arXiv:1411.6272v2* (2014).
- [17] HERMAN, M., AND STROHMER, T. High resolution radar via compressed sensing. *IEEE Trans. Signal Process.* 57, 6 (2009), 2275–2284.
- [18] HERMAN, M., AND STROHMER, T. General deviants: An analysis of perturbations in compressed sensing. *Selected Topics in Signal Processing, IEEE Journal of* 4, 2 (April 2010), 342–349.
- [19] HÜGEL, M., RAUHUT, H., AND STROHMER, T. Remote sensing via  $\ell_1$ -minimization. *Found. Comput. Math.* 14 (2014), 115–150.
- [20] KRAHMER, F., MENDELSON, S., AND RAUHUT, H. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.* 67, 11 (2014), 1877–1904.
- [21] LI, J., AND STOICA, P. MIMO Radar with Colocated Antennas. *Signal Processing Magazine, IEEE* 24, 5 (Sept 2007), 106–114.

- [22] LI, J., AND STOICA, P., Eds. *MIMO Radar Signal Processing*. John Wiley & Sons, 2009.
- [23] LUST PIQUARD, F. Inégalités de Khintchine dans  $c_p(1 < p < \infty)$ . *C. R. Acad. Sci. Paris S'ér. I Math.* 303 (1986), 289–292.
- [24] PISIER, G., AND LUST PIQUARD, F. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.* 29, 2 (1991), 241–260.
- [25] RAUHUT, H. Compressive sensing and structured random matrices. In *Theoretical foundations and numerical methods for sparse recovery*, M. Fornasier, Ed., vol. 9 of *Radon Series Comp. Appl. Math.* deGruyter, 2010, pp. 1–92.
- [26] STROHMER, T., AND FRIEDLANDER, B. Compressed sensing for MIMO radar - algorithms and performance. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on* (Nov 2009), pp. 464–468.
- [27] STROHMER, T., AND FRIEDLANDER, B. Analysis of Sparse MIMO Radar. *Appl. Comput. Harm. Anal.* 37 (2014), 361–388.
- [28] TALAGRAND, M. *The Generic Chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.
- [29] TALAGRAND, M. *Upper and lower bounds for stochastic processes*, vol. 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014.
- [30] TROPP, J. A. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* 12, 4 (2012), 389–434.
- [31] YU, Y., PETROPULU, A., AND POOR, H. Measurement Matrix Design for Compressive Sensing-Based MIMO Radar. *Signal Processing, IEEE Transactions on* 59, 11 (Nov 2011), 5338–5352.
- [32] YU, Y., PETROPULU, A., AND POOR, H. CSSF MIMO RADAR: Compressive-Sensing and Step-Frequency Based MIMO Radar. *Aerospace and Electronic Systems, IEEE Transactions on* 48, 2 (April 2012), 1490–1504.